Available online at www.sciencedirect.com

SCIENCE DIRECT®

Computers & Operations Research 32 (2005) 2523–2541

computers & operations research

www.elsevier.com/locate/dsw

A novel nonlinear ensemble forecasting model incorporating GLAR and ANN for foreign exchange rates☆

Lean Yu[a,b], Shouyang Wang[a,1], K.K. Lai[c,*]

[a]Institute of Systems Science, Academy of Mathematics and Systems Sciences, Chinese Academy of Sciences, Beijing 100080, China

[b]Graduate School of Chinese Academy of Sciences, Beijing 100080, China

[c]Department of Management Sciences, City University of Hong Kong, Tat Chee Avenue, Kowloon, Hong Kong, China

Available online 29 July 2004

Abstract

In this study, we propose a novel nonlinear ensemble forecasting model integrating generalized linear auto-regression (GLAR) with artificial neural networks (ANN) in order to obtain accurate prediction results and ameliorate forecasting performances. We compare the new model's performance with the two individual forecasting models—GLAR and ANN—as well as with the hybrid model and the linear combination models. Empirical results obtained reveal that the prediction using the nonlinear ensemble model is generally better than those obtained using the other models presented in this study in terms of the same evaluation measurements. Our findings reveal that the nonlinear ensemble model proposed here can be used as an alternative forecasting tool for exchange rates to achieve greater forecasting accuracy and improve prediction quality further.

Keywords: GLAR; ANN; Nonlinear ensemble model; Forecasting; Foreign exchange rates

1. Introduction

Exchange rate modeling and forecasting has been a common research stream in the last few decades. Over this time, the research stream has gained momentum with the advancement of computer

☆ Partially supported by NSFC, CAS, JSPS and strategic research grant of City University of Hong Kong.
* Corresponding author. Tel.: 852-27888563; fax: 852-27888560.
E-mail address: mskklai@cityu.edu.hk (K.K. Lai).
[1] Also with the Institute of Policy and Planning, University of Tsukuba, Japan.

0305-0548/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.cor.2004.06.024

technologies, which have made many elaborate computation methods available and practical. However, it is not easy to predict exchange rates due to their high volatility and noise. But the difficulty in forecasting exchange rates is usually attributed to the limitation of many conventional forecasting models; this has encouraged academic researchers and business practitioners to develop more predictable forecasting models. As a result models using artificial intelligence such as artificial neural network (ANN) techniques have been recognized as more useful than conventional statistical forecasting models. Literature documenting the research shows this is quite diverse and involves different architectural designs. Some examples are presented. De Matos [1] compared the strength of a multilayer feed-forward network (MLFN) with that of a recurrent network based on the forecasting of Japanese yen futures. Kuan and Liu [2] provided a comparative evaluation of the performance of MLFN and a recurrent network on the prediction of an array of commonly traded exchange rates. Hsu et al. [3] developed a clustering neural network model to predict the direction of movements in the USD/DEM exchange rate. Their experimental results suggested that their proposed model achieved better forecasting performance relative to other indicators.

Similarly, Tenti [4] applied recurrent neural network models to forecast exchange rates. Shazly and Shazly [5] designed a hybrid model combining neural networks and genetic training to the 3-month spot rate of exchange for four currencies: the British pound, the German mark, the Japanese yen and the Swiss franc. The experimental results reported revealed that the networks' forecasts outperformed predictions made by both the forward and futures rates in terms of accuracy and correctness. In a more recent study by Leung et al. [6], the forecasting accuracy of MLFN was compared with the general regression neural network (GRNN). The study showed that the GRNN possessed a greater forecasting strength relative to MLFN with respect to a variety of currency exchanges. Zhang and Berardi [7] adopted a different approach. Instead of using single network architecture, their research investigated the use of ensemble methods in exchange rate forecasting. Essentially, the study proposed using systematic and serial partitioning methods to build ensemble models consisting of different neural network structures. Results indicated that the ensemble network could consistently outperform a single network design.

Recently, more hybrid forecasting models have been developed that integrate neural network techniques with many conventional and burgeoning forecasting methods such as econometrical models and time series models to improve prediction accuracy. There are a few examples in the existing literature combining neural network forecasting models with conventional time series forecasting techniques, especially auto-regression integrated moving average (ARIMA). Wedding II and Cios [8] constructed a combination model incorporating radial basis function neural networks (RBFNN) and the univariant Box-Jenkins (UBJ) model to predict the Wolfer Sunspot, West German monthly unemployment figures and the number of monthly housing starts; the experiments proved that the hybrid model for time series forecasting is capable of giving significantly better results. Luxhoj et al. [9] presented a hybrid econometric and ANN approach for sales forecasting and obtained good prediction performance. Likewise, Voort et al. [10] introduced a hybrid method called KARIMA using a Kohonen self-organizing map and ARIMA method to predict short-term traffic flow, and obtained relatively better results.

In the same way, Tseng et al. [11] proposed using a hybrid model (SARIMABP) that combines the seasonal ARIMA (SARIMA) model and the back-propagation (BP) neural network model to predict seasonal time series data. The experimental results showed that SARIMABP was superior to the SARIMA model, to the back-propagation model with deseasonalized data, and to the back-propagation model with differenced data for the test cases of machinery production time series and soft drink time series. The values of the MSE, MAE and MAPE were all the lowest for the SARIMABP model. The SARIMABP also outperformed other models in terms of overall proposed criteria including MSE, MAE, MAPE and turning points forecasts. For the machinery production time series, the SARIMABP model remained stable even when the number of historical data was reduced from 60 to 36. Zhang [12] proposed a hybrid methodology that combined both ARIMA and ANN models taking advantage of the unique strengths of these models in linear and nonlinear modeling for time series forecasting. Empirical results with real data sets indicated that the hybrid model could provide an effective way to improve the forecasting accuracy achieved by either of the models used separately.

In their pioneering work on combined forecasts [13,14], Bates and Granger showed that a linear combination of forecasts would give a smaller error variance than any of the individual methods. Since then, the studies on this topic (i.e. combined forecasts) have expanded dramatically. Makridakis et al. [15] claimed that using a hybrid model or combining several models has become common practice in improving forecasting accuracy ever since the well-known M-competition in which a combination of forecasts from more than one model often leads to improved forecasting performance. Likewise, Pelikan et al. [16] and Ginzburg and Horn [17] proposed combining several feed-forward neural networks to improve time series forecasting accuracy. In 1989, Clemen [18] provided a comprehensive review and annotated bibliography in this area. The basic idea of the model combination in forecasting is to use each model's unique feature to capture different patterns in the data. Both theoretical and empirical findings suggest that combining different methods can be an effective and efficient way to improve forecast performances.

However, we find that there are two main problems in the existing hybrid models and combined models mentioned above. In this literature, both the hybrid forecasting models and combined forecasting models are limited to linear combination form. A linear combination approach is not necessarily appropriate for all the circumstances (Problem I). Furthermore, it is not easy to determine the number of individual forecasting models. It is well known to us that not all circumstances are satisfied with the rule of "the more, the best". Thus, it is necessary to choose an appropriate method to determine the number of individual models for combining forecasting (Problem II).

In view of the first problem (i.e. linear combination drawback) a nonlinear ensemble forecasting model utilizing the ANN technique is introduced in this paper. The fact that we use "ensemble" rather than "combined" is based on semantics. According to the American Heritage Dictionary [19], "ensemble" is defined as "a unit or group of complementary parts that contribute to a single effect", while "combine" is "to bring into a state of unity; or merge". In fact, the nonlinear ensemble model is equivalent to the nonlinear combination model in the general sense. But the word "ensemble" is preferable to "combined" from the semantic point of view.

In view of the second problem (i.e. determining the number of individual models for combining forecasting), the principal component analysis (PCA) technique is introduced. We use the PCA technique to choose the number of individual forecasting models.

Considering the previous two main problems, this study proposes a novel nonlinear ensemble forecasting model for exchange rate forecasting. This model utilizes the ANN technique and PCA technique, integrates GLAR with ANN models, and takes full advantage of hybrid methods and combined techniques. The proposed novel nonlinear ensemble model is the PCA&ANN-based nonlinear ensemble forecasting approach, which integrates GLAR and ANN models as well as the single hybrid methodology. The aims of this study are two-fold: (1) to show how to predict exchange rates using the proposed nonlinear ensemble model and (2) to display how various methods compare in their accuracy in forecasting foreign exchange rates. In view of the two aims, this study mainly describes the building process of the proposed nonlinear ensemble model and the application of the nonlinear ensemble forecasting approach in foreign exchange rate forecasting between the US dollar and three other major currencies—German marks, British pound and Japanese yen—incorporating GLAR and ANN, while comparing forecasting performance with all kinds of evaluation criteria.

The rest of the study is organized as follows. The next section describes the model building process in detail. In order to verify the effectiveness and efficiency of the proposed model, empirical analysis of the three main currencies' exchange rates is reported in Section 3. The conclusions are contained in Section 4.

2. Model building processes

In this section, we mainly describe the model building process step by step and in detail. First we present the GLAR and ANN modeling processes briefly. Then a hybrid forecasting method and combined forecasting methodologies are introduced with the aid of existing literature. Finally, based on previous works, a nonlinear ensemble model is proposed and three forecasting evaluation criteria are presented.

2.1. Generalized linear auto-regression (GLAR) model

The generalized linear auto-regression (GLAR) model, first introduced by Shephard [20], is equivalent to a system of reduced form equations relating each endogenous variable to lag endogenous (predetermined) and pertinent exogenous variables. That is, in a GLAR model, the future value of a variable is assumed to be a linear function of several past observations and random errors. Generally, the form of the GLAR model is given in the following.

$$y_t = \alpha + C(L)y_{t-1} + D(L)x_{t-1} + \varepsilon_t, \tag{1}$$

where y_t and ε_t are the actual value of endogenous variables and random disturbance at time t, x_{t-1} is the actual value of related exogenous variables, α is a constant term, $C(L)$ and $D(L)$ are the pth and qth order lag polynomial of endogenous variable and exogenous variables respectively in the lag operator L, such that $C(L) = C_0 - C_1L - C_2L^2 - \cdots - C_pL^P$ and $D(L) = D_0 - D_1L - D_2L^2 - \cdots - D_qL^q$. L denotes the lag operator, e.g., $Ly_t = y_{t-1}$, $L^2y_t = y_{t-2}$ and so on, and random disturbances, ε_t are assumed to be independently and identically distributed with a mean of zero and a constant variance of σ^2, i.e. $\varepsilon_t \sim \mathrm{IID}(0, \sigma^2)$.

Actually, the GLAR is a special form of vector auto-regression (VAR) model [21] by comparison. The difference between GLAR and VAR is only in the representation of variables. That is, the former is a form of individual variable and the latter is the vector form. Similarly, the fact that we choose the GLAR model rather than autoregressive integrated moving average (ARIMA) model [22] is based on the following idea: the ARIMA model does not contain the term of exogenous variables (x_t) and thus cannot capture the effect of other external factors. Furthermore, adding an explanatory variable may strengthen the model's forecasting capability. Therefore, from this paper's perspective, because foreign exchange rates are often affected by many factors and all variables are time series, it seems that the GLAR model is simpler and easier to understand than the VAR model and ARIMA model, and hence the GLAR model is preferable to VAR and ARIMA.

Based on earlier works [20,23,24], the GLAR model involves the following five-step iterative procedures:

(a) Stationary test of time series, such as unit root test;
(b) Identification of the GLAR structure, i.e. the model order is determined;
(c) Estimation of the unknown parameters;
(d) Model checks and diagnostics;
(e) Forecast future outcomes based on the known data.

This five-step model building process is typically repeated several times until a satisfactory model is finally selected. The final model selected can then be used for prediction purposes.

The advantage of GLAR is that it is capable of receiving external information from exogenous variables. For example, the foreign exchange rate is often influenced by many external factors (e.g. exports, inflation and interest rate, etc.) besides the effect of formation mechanism of itself. Furthermore, the GLAR model fits the linear characteristic of time series well. The disadvantage of the GLAR is that it cannot capture nonlinear patterns of financial time series if nonlinearity exists.

2.2. *Artificial neural network (ANN) model*

In this study, one of the widely used ANN models, the back-propagation neural network (BPNN), is used for time series forecasting [12,25–28]. Usually, the BPNN model consists of an input layer, an output layer and one or more intervening layers also referred to as hidden layers. The hidden layers can capture the nonlinear relationship between variables. Each layer consists of multiple neurons that are connected to neurons in adjacent layers. Since these networks contain many interacting nonlinear neurons in multiple layers, the networks can capture relatively complex phenomena. ANNs are already one of the models that is able to approximate various nonlinearities in the data series.

A neural network can be trained by the historical data of a time series in order to capture the nonlinear characteristics of the specific time series. The model parameters (connection weights and node biases) will be adjusted iteratively by a process of minimizing the forecasting errors. For time series forecasting, the final computational form of the ANN model is as

$$y_t = a_0 + \sum_{j=1}^{q} w_j f\left(a_j + \sum_{i=1}^{p} w_{ij} y_{t-i}\right) + \xi_t, \tag{2}$$

where a_j $(j = 0, 1, 2, \ldots, q)$ is a bias on the jth unit, and w_{ij} $(i = 1, 2, \ldots, p;\ \ j = 1, 2, \ldots, q)$ is the connection weight between layers of the model, $f(\cdot)$ is the transfer function of the hidden layer, p is the number of input nodes and q is the number of hidden nodes. Actually, the ANN model in (2) performs a nonlinear functional mapping from the past observation $(y_{t-1}, y_{t-2}, \ldots, y_{t-p})$ to the future value y_t, i.e.,

$$y_t = \varphi(y_{t-1}, y_{t-2}, \ldots, y_{t-p}, v) + \xi_t, \tag{3}$$

where v is a vector of all parameters and φ is a function determined by the network structure and connection weights. Thus, in some senses, the ANN model is equivalent to a nonlinear autoregressive (NAR) model.

It is worth noting that in this paper the ANN model is the typical back-propagation neural network; however, the core-training algorithm is not the gradient descent rule, but the Levenberg–Marquardt rule, as this algorithm can increase the network's training speed. The Levenberg–Marquardt rule is a kind of quick algorithm in which the network's matrix of weights is updated based on the Jacobian matrix, J, collecting the partial derivatives of the network error e with respect to the weights. In other words, the matrix ΔW collecting the corrections of the weights in matrix W is computed according to

$$\Delta W = (J^T J + \mu I)^{-1} J^T e. \tag{4}$$

Here μ is a constant. If μ is sufficiently large, the previous algorithm is similar to the gradient descent algorithm; if μ is equal to zero, the previous rule will be a Gauss–Newton algorithm. The advantage of the Levenberg–Marquardt algorithm lies in the fact that the Hessian matrix is not needed; moreover using the Levenberg–Marquardt algorithm can accelerate the network's training speed, save training time and achieve better training results simultaneously.

The processes for developing the back-propagation neural network model are as follows (these are as proposed by Ginzburg and Horn [17]):

(a) normalize the learning set;
(b) decide the architecture and parameters: i.e., learning rate, momentum, and architecture. There are no criteria in deciding the parameters other than a trial-and-error basis;
(c) initialize all weights randomly;
(d) Training, where the stopping criterion is either the number of iterations reached or when the total sum of squares of error is lower than a predetermined value;
(e) choose the network with the minimum error (i.e., model selection criterion);
(f) forecast future outcome.

Readers interested in a more detailed introduction to ANN modeling are referred to the related literature [17,25–28].

The major advantage of ANN models is their flexible nonlinear modeling capability. They can capture the nonlinear characteristics of time series well. However, using ANN to model linear problems may produce mixed results [12,29,30]. Therefore, we can conclude that the relationship between GLAR and ANN is complementary. To take full advantage of the individual strengths of two models, it is necessary to integrate the GLAR and ANN models, as mentioned earlier.

2.3. The hybrid methodology integrated GLAR with ANN

In real life, a time series forecasting problem such as exchange rate prediction is far from simple due to high volatility, complexity, irregularity and noisy market environment. Furthermore, real-world time series are rarely pure linear or nonlinear. They often contain both linear and nonlinear patterns. If this is the case, there is no omnipotent model that is suitable for all kinds of time series data. Although both GLAR and ANN models have achieved success in their own linear or nonlinear domains, neither GLAR nor ANN can adequately model and predict time series since the linear models cannot deal with nonlinear relationships while the ANN model alone is not able to handle both linear and nonlinear patterns equally well [12]. On the other hand, as previously mentioned, for time series forecasting the relationship between GLAR and ANN is complementary. GLAR is a class of linear models that can capture time series' linear characteristics, while ANN models trained by back-propagation with hidden layers are a class of general function approximators capable of modeling nonlinearity and which can capture nonlinear patterns in time series. Commixing the two models may yield a robust method, and more satisfactory forecasting results may be obtained by incorporating a time series model and an ANN model. Therefore, we propose a hybrid model integrating GLAR and ANN for exchange rate forecasting. In view of the works of [12,31], the proposed hybrid model is a two-phase forecasting procedure, which incorporating GLAR model with ANN model in an adaptive (sequential) manner. The motivation of hybrid methodology is to create a synergy effect that further improves the prediction power.

In the first phase, a GLAR model is used to fit the linear component of time series, which is assumed to be $\{y_t, t=1, 2, \ldots\}$, and generate a series of forecasts, which is defined as $\{\hat{L}_t\}$. However, since some time series, such as exchange rates, often contain more complex patterns than those of linear regression models in the practical application, it is not enough to fit only a linear component for a time series. Hence, we also need to discover the nonlinear relationship of the time series in order to improve prediction accuracy. By comparing the actual value y_t of the time series and forecast value $\hat{L}_t$ of the linear component, we can obtain a series of nonlinear components, which is assumed to be $\{e_t\}$. That is,

$$e_t = y_t - \hat{L}_t. \tag{5}$$

Thus, a nonlinear time series is obtained. The next step is how to fit the nonlinear component of series.

In the second phase of the hybrid methodology, a neural network model is used to model the above nonlinear time series. By training the ANN model using previously generated nonlinear time series as inputs, the trained ANN model is then used to generate a series of forecasts of nonlinear components of time series, defined by $\{\hat{N}_t\}$.

In order to obtain the synergetic forecast results, the final forecasting results, defined by $\{\hat{y}_t\}$, are calculated as

$$\hat{y}_t = \hat{L}_t + \hat{N}_t. \tag{6}$$

In summary, the proposed hybrid methodology consists of two phases or four steps. In the first step, a GLAR model should be identified and the corresponding parameters should be estimated, i.e., a GLAR model is constructed. In the second step, the nonlinear components are computed from the GLAR model. In the third step, a neural network model is developed to model the nonlinear components. In the final step, the combined forecast results are obtained from Eq. (6).

Theoretically speaking, the hybrid model presented here is still a class of single forecasting model. However, many studies reveal that combining several different forecasts can further improve prediction performances [13–16,32–34]. In the following, we introduce the combined forecasting model briefly and then point out some problems with them.

2.4. *The combined forecasting model*

The idea of using a combined forecasting model is not new. Since the pioneering works of Reid [13] and of Bates and Granger [14], a variety of studies have been conducted which showed that a combination of forecasts often outperforms the forecasts obtained from a single method. The main problem of combined forecasts can be described as follows. Suppose there are n forecasts such as $\hat{y}_1(t), \hat{y}_2(t), \ldots, \hat{y}_n(t)$. The question is how to combine (ensemble) these different forecasts into a single forecast $\hat{y}(t)$, which is assumed to be a more accurate forecast. The general form of the model for such a combined forecast can be defined as

$$\hat{y}(t) = \sum_{i=1}^{n} w_i \hat{y}_i(t), \tag{7}$$

where w_i denotes the assigned weight of $\hat{y}(t)$, and in general the sum of the weights is equal to 1, i.e. $\sum_i w_i = 1$. In some situations the weights may have negative values and the sum of them may be greater than one [34].

There are a variety of methods available to determine the weights used in the combined forecasts. First of all, the equal weights (EW) method, which uses an arithmetic average of the individual forecasts, is a relatively easy and robust method. In this method,

$$w_i = \frac{1}{n}, \tag{8}$$

where n is the number of forecasts. However, since the weights in the combination are so unstable, in practice a simple average may not be the best technique to use [33]. Moreover, research indicates that this weighted method has not always performed well empirically [34]. A new method, called the minimum-error (ME) method, is introduced in this paper.

The ME method is an approach that minimizes the forecasting errors when combining individual forecasts into a single forecast. We propose a new solution for this method using linear programming (LP). Its principle and computational process are as follows.

Set the sum of absolute forecasting error $\left(\text{i.e., } \sum_i e_i(t) \text{ at time } t\right)$ as

$$Q = \sum_{i=1}^{n} |e_i(t)| = \sum_{i=1}^{n} |w_i(t)(\hat{y}_i(t) - y_i(t))| \quad (t = 1, 2, \ldots, N). \tag{9}$$

That is, Q is the objective function of LP. In order to eliminate the absolute sign of objective function, assume that

$$\begin{aligned} u_i(t) &= \frac{|e_i(t)| + e_i(t)}{2} = \begin{cases} e_i(t), & e_i(t) \geqslant 0 \\ 0, & e_i(t) < 0 \end{cases}, \quad v_i(t) = \frac{|e_i(t)| - e_i(t)}{2} \\ &= \begin{cases} 0, & e_i(t) \geqslant 0 \\ -e_i(t), & e_i(t) < 0 \end{cases} \ldots \end{aligned} \tag{10}$$

Note that the aim of introducing $u_i(t)$ and $v_i(t)$ is to transform the absolute sign of the objective function so as to be consistent with the standard form of linear programming. Clearly, $|e_i(t)| = u_i(t) + v_i(t)$,

$e_i(t) = u_i(t) - v_i(t)$. Based on the above specification we can construct the linear programming model as follows:

$$\text{(LP)} \begin{cases} Min\, Q = \sum_{i=1}^{n} (u_i(t) + v_i(t)), \\ \sum_{i=1}^{n} [\hat{y}_i(t) - y_i(t)] w_i(t) - u_i(t) + v_i(t) = 0, \\ \sum_{i=1}^{n} w_i(t) = 1, \\ w_i(t) \geqslant 0, \quad u_i(t) \geqslant 0, \quad v_i(t) \geqslant 0, \quad i = 1, 2, \ldots, n, \quad t = 1, 2, \ldots, N, \end{cases} \tag{11}$$

where i denotes the number of individual forecasting models, and t represents the forecasting periods.

In Equation group (11), the aim of assuming $w_i \geqslant 0$ is to make every forecast method contribute to the combined forecasting results. The ME method is equivalent to a simple dynamic linear programming problem; thus, the optimal solutions are solved by the simplex algorithm. In fact, the ME method is equivalent to a time-variant weight combination methodology.

However, there are two main problems in existing combined forecasting approaches. One is the number of individual forecasts. Theoretical proof of [35] shows that the total forecasting errors of combined forecasts does not necessarily decrease with an increase of the number of individual forecasting models. If this is the case, there are redundant or repeated (even noisy) information among the individual forecasts. Furthermore, the redundant information often directly impacts on the effectiveness of the combined forecasting method. It is therefore necessary to eliminate some individual forecasts that are not required. In other words, key problem of combined forecasts is to determine a reasonable number of individual forecasts. The other problem is that the relationships among individual forecasts are determined in a linear manner. The two methods mentioned above and many other weighted methods [33] are all developed under linear consideration. A combined forecast should merge the individual forecasts according to the natural relationships existing between the individual forecasts, including but not limited to linear relationships. Because a linear combination of existing information cannot represent the relationship among individual forecasting models in some situations, it is necessary to introduce nonlinear combination methodology in the combined forecasts. Subsequently, we set out to solve the two problems in the following subsection.

2.5. The nonlinear ensemble (NE) forecasting model

In view of the two main problems outlined in the previous subsection, a novel nonlinear ensemble model is proposed. The following is the process of building the proposed model for the two problems.

In the first problem, the question we face is how to extract effective information that reflects substantial characteristics of series from all selected individual forecasts and how to eliminate redundant information. The PCA technique (see [36,37]) is used as an alternative tool to solve the problems. The PCA technique, an effective feature extraction method, is widely used in signal processing, statistics and neural computing. The basic idea in PCA is to find the components $(s_1, s_2, \ldots, s_p)$ that can explain the maximum amount of variance possible by p linearly transformed components from data vector with q dimensions. The mathematical technique used in PCA is called eigenanalysis. In addition, the basic goal in PCA is to reduce the dimension of the data. Thus, one usually chooses $p \leqslant q$. Indeed, it can be proven that the representation given by PCA is an optimal linear dimension reduction technique in the mean-square sense [36]. Such a reduction in dimension has important benefits. First, the computation of the subsequent processing is reduced. Second, noise may be reduced and the meaningful underlying information identified. The following presents the PCA process for combined forecasting.

Assuming that there are n individual forecasting models in the combined forecasts and that every forecasting model contains m forecasting results, then forecasting matrix (Y) can represented as

$$Y = \begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1m} \\ y_{21} & y_{22} & \cdots & y_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ y_{n1} & y_{n2} & \cdots & y_{nm} \end{bmatrix}, \tag{12}$$

where y_{ij} is the jth forecasting value with the ith forecasting model.

Next, we deal with the forecasting matrix using the PCA technique. First, eigenvalues $(\lambda_1, \lambda_2, \ldots, \lambda_n)$, and corresponding eigenvectors $A = (a_1, a_2, \ldots, a_n)$ can be solved from the forecasting matrix. Then the new principal components are calculated as

$$Z_i = a_i^{\mathrm{T}} Y \quad (i = 1, 2, \ldots, n). \tag{13}$$

Subsequently, we choose m $(m = n)$ principal components from existing n components. If this is the case, the saved information content is judged by

$$\theta = (\lambda_1 + \lambda_2 + \cdots + \lambda_m)/(\lambda_1 + \lambda_2 + \cdots + \lambda_n). \tag{14}$$

If θ is sufficiently large (e.g., $\theta > 0.8$), enough information has been saved after the feature extraction process of PCA. Thus, re-combining the new information can further improve the prediction performance of combined forecasts.

For further explanation, a simple example is presented. Assuming that we predict the GBP/USD exchange rate using three different forecasting methods (such as GLAR, ANN and ARIMA) and obtain eight forecasts for every method, the corresponding forecasting matrix is shown below.

Original forecasting matrix:

$$Y = \begin{bmatrix} 0.6723 & 0.6599 & 0.6474 & 0.6349 & 0.6224 & 0.6099 & 0.5974 & 0.5848 \\ 0.6712 & 0.6586 & 0.6471 & 0.6356 & 0.6251 & 0.6155 & 0.6064 & 0.5982 \\ 0.6697 & 0.6566 & 0.6436 & 0.6310 & 0.6186 & 0.6064 & 0.5946 & 0.5829 \end{bmatrix}.$$

Before using combined forecasting methodology, the PCA is performed. Accordingly, its eigenvalues and eigenvectors are computed as

$$\lambda = \begin{pmatrix} 0.1762E-01 \\ 0.1288E-04 \\ 0.9019E-08 \end{pmatrix}, \quad A = \begin{pmatrix} -0.5806 & 0.5390 & 0.6102 \\ 0.8084 & 0.2928 & 0.5106 \\ 0.0966 & -0.7898 & 0.6058 \end{pmatrix}.$$

According to the previous computation process, the new forecasting matrix (or new principal components) can be as

New forecasting matrix:

$$Y' = \begin{bmatrix} 0.0734 & 0.0514 & 0.0301 & 0.0089 & 0.0115 & 0.0315 & 0.0510 & 0.0699 \\ 0.0019 & 0.0002 & 0.0006 & 0.0014 & 0.0014 & 0.0008 & 0.0003 & 0.0020 \\ 0.0000 & 0.0000 & 0.0001 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \end{bmatrix}.$$

From the information presented in the new forecasting matrix, we can find and conclude that the third row belongs to redundancy and should be eliminated, indicating that the third individual forecasting model provides repeated information and cannot improve combined forecasting performance: thus this forecasting method should be removed from the combined forecasts. Therefore, the reasonable number of individual forecasting methods is determined. Of course, the value of θ should be calculated for judgment if the new forecasting matrix is not obvious. In this example, $\theta = (0.1762E-01+0.1288E-04)/(0.1762E-01+0.1288E-04+0.9019E-08) = 0.999999489$. This also shows that the former two methods contain enough information to combine forecasts.

In the case of the second problem, we propose a nonlinear ensemble forecasting model as a remedy. The detailed model is presented as follows.

A nonlinear ensemble forecasting model can be viewed as nonlinear information processing system which can be represented as

$$y = f(I_1, I_2, \ldots, I_n), \tag{15}$$

where $f(\cdot)$ is a nonlinear function and I_i denotes the information provided by individual forecasting models. If individual forecasting model can provide an individual forecast $\hat{y}_i$, then Eq. (15) can be represented as

$$\hat{y} = f(\hat{y}_1, \hat{y}_2, \ldots, \hat{y}_n). \tag{16}$$

Determining the function $f(\cdot)$ is quite challenging. In this study, ANN is employed to realize nonlinear mapping. The ability of back-propagation neural networks to represent nonlinear models has been tested by previous work.

In fact, the ANN training is a process of searching for optimal weights. That is, this training process made the sum of the square errors minimal, i.e.,

$$Min[(y-\hat{y})(y-\hat{y})^{\mathrm{T}}] \tag{17}$$

or

$$Min\{[y - f(\hat{y}_1, \hat{y}_2, \ldots, \hat{y}_n)][y - f(\hat{y}_1, \hat{y}_2, \ldots, \hat{y}_n)]^{\mathrm{T}}\}. \tag{18}$$

By solving the earlier two problems of linear combination models, a PCA&ANN-based nonlinear ensemble model is generated for nonlinear combining forecasting.

To summarize, the proposed nonlinear ensemble model consists of four stages. Generally speaking, in the first stage we construct an original forecasting matrix according to the selected individual forecasts. In the second stage, the PCA technique is used to deal with the original forecasting matrix and a new forecasting matrix is obtained. In the third stage, based upon the judgments of PCA results, the number of individual forecasting models is determined. And in the final stage, an ANN model is developed to ensemble different individual forecasts, meantime the corresponding forecasting results are obtained.

After completing the proposed nonlinear ensemble model, we want to know whether the proposed model indeed improves forecasting accuracy or not. In this study, we use an individual GLAR model, an individual ANN model, the hybrid model, the linear combination models, and the PCA&ANN-based nonlinear ensemble model to predict exchange rates so as to compare forecasting performance. The basic flow diagram is shown in Fig. 1.

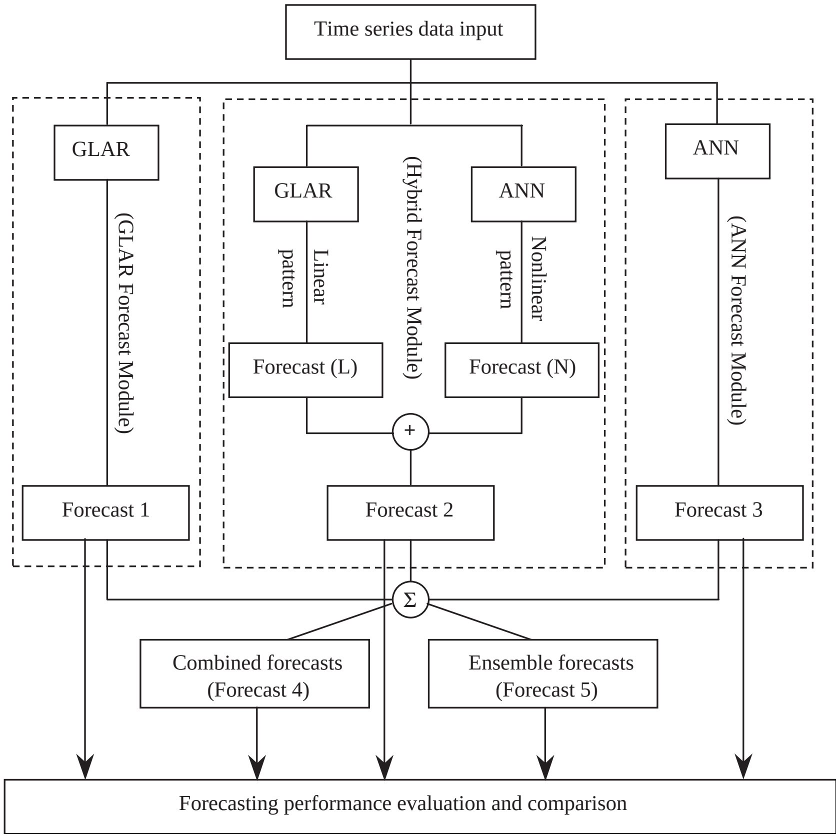

Fig. 1. A flow diagram of the combined and nonlinear ensemble forecast.

2.6. Forecasting evaluation criteria

As Weigend [38] argued, a measure normally used to evaluate and compare the predictive power of the model is the normalized mean squared error (NMSE) (this was used to evaluate entries in the Santa Fe Time Series Competition). Given N pairs of the actual values (or targets, y_t) and predicted values $(\hat{y}_t)$, the NMSE which normalizes the *MSE* by dividing it through the variance of respective series can be defined as

$$\mathrm{NMSE} = \frac{\sum_{t=1}^{N}(y_t - \hat{y}_t)^2}{\sum_{t=1}^{N}(y_t - \bar{y}_t)^2} = \frac{1}{\sigma^2} \cdot \frac{1}{N} \cdot \sum_{t=1}^{N}(y_t - \hat{y}_t)^2, \tag{19}$$

where σ^2 is the estimated variance of the data and $\bar{y}_t$ the mean.

Clearly, accuracy is one of the most important criteria for forecasting models—the others being the cost savings and profit earnings generated from improved decisions. From the business point of view, the latter is more important than the former. For business practitioners, the aim of forecasting is to support or improve decisions so as to make more money. Thus profits or returns are more important than conventional fit measurements. But in exchange rate forecasting, improved decisions often depend on correct forecasting directions or turning points between the actual and predicted values, y_t and $\hat{y}_t$, respectively, in the testing set with respect to directional change of exchange rate movement (expressed in percentages). The ability to forecast movement direction or turning points can be measured by a statistic developed by Yao and Tan [39]. Directional change statistics (D_{stat}) can be expressed as

$$D_{\text{stat}} = \frac{1}{N}\sum_{t=1}^{N} a_t \times 100\%, \tag{20}$$

where $a_t = 1$ if $(y_{t+1} - y_t)(\hat{y}_{t+1} - y_t) \geqslant 0$, and $a_t = 0$ otherwise.

However, the real aim of forecasting is to obtain profits based on prediction results. Here the annual return rate is introduced as another evaluation criterion. Without considering the friction costs, the annual return rate is calculated according to the compound interest principle.

$$R = \left(\left(\frac{M}{S}\right)^{12/N} - 1\right) \times 100, \tag{21}$$

where R is the annual return rate, M is the amount of the money obtained on the last month of testing, S is the amount of the money used for trading on the first month of testing, and N is the number of months in the testing periods. It is worth noting that the computation of R is based on the trading strategy in the following:

$$\text{If } (\hat{y}_{t+1} - y_t) > 0, \text{ then `` buy''}, \text{ else `` sell''}, \tag{22}$$

where y_t is the actual value at time t, $\hat{y}_{t+1}$, is the prediction at time $t+1$. That is, we use the difference between the predicted value and the actual value to guide trading.

We want to use the latter two statistics as the main evaluation criteria because the normalized *MSE* measure predictions only in terms of levels. Hence, it is more reasonable to choose D_{stat} and annual return rate (R) as the measurements of forecast evaluation. Of course, NMSE is also taken into consideration in terms of a comparison of levels.

3. Empirical analysis

3.1. Data description

The foreign exchange data used in this paper are monthly and are obtained from Pacific Exchange Rate Service (http://fx.sauder.ubc.ca/), provided by Professor Werner Antweiler, University of British Columbia, Vancouver, Canada. They consist of the US dollar exchange rate against each of the currencies (DEM, GBP and JPY) studied in this paper. We take monthly data from January 1971 to December 2000 as in-sample (training periods) data sets (360 observations including 24 samples for validation). We also take the data from January 2001 to December 2003 as out-of-sample (testing periods) data sets (36 observations) which is used to evaluate the good or bad performance of prediction based on some evaluation measurement. For exogenous variables' consideration of GLAR, we can choose exports, inflation rate, interest rates and other factors related to exchange rates as exogenous variables. Unfortunately,

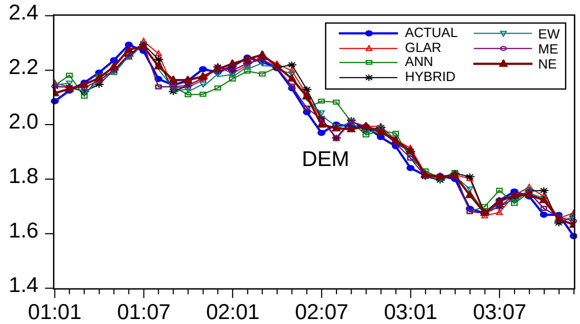

Fig. 2. A graphical comparison of DEM rate prediction results using different models (2001:01–2003:12).

monthly inflation rate and interest rate data is hard to obtain. Therefore, we only choose "exports" as the exogenous variable for the GLAR model. The monthly export data in this paper is taken directly from the CD-ROM of International Financial Statistics of IMF. In addition, all the evaluations in this paper (with the exception of specifications) are based on the whole testing sets. In order to save space, the original data are not listed here, and detailed data can be obtained from the website or from the authors.

3.2. *Empirical results*

In this study, all GLAR models are implemented via the Eviews software package, which is produced by Quantitative Micro Software Corporation. The individual ANN model, the linear combination model with minimum-error method and nonlinear ensemble model are built using the Matlab software package, which is produced by Mathworks Laboratory Corporation. The single hybrid methodology involves the GLAR and ANN models and thus it use both the software packages. The ANN models use trial and error to determine the network architecture of 4-4-1 by minimizing the forecasting error with the exception of specifications. According to the idea shown in Fig. 1, the single GLAR model, the single ANN model, the single hybrid model, and the linear combination model with equal weights method, the linear combination model with minimum-error method, and the PCA&ANN-based nonlinear ensemble model are used to predict the exchange rates of three main currencies for comparison. For space reasons the computational processes are omitted but can be obtained from the authors if required.

Figs. 2–4 give graphical representations of the forecasting results for exchange rates of the three main currencies using different models. Tables 1–3 show the forecasting performance of different models from different perspectives. From the graphs and tables, we can generally see that the forecasting results are very promising for all currencies under study either where the measurement of forecasting performance is goodness of fit such as NMSE (refer to Table 1) or where the forecasting performance criterion is D_{stat} (refer to Table 2).

In detail, Fig. 2 shows the comparison for the DEM rate of the nonlinear ensemble forecasting results versus the forecast results of individual models, the hybrid model, and linear combination model. Similarly, it can be seen from Fig. 3 that forecast accuracy for the GBP rate has significantly improved using the nonlinear ensemble forecasting models. Fig. 4 provides the comparison of the JPY prediction results using different models. The results indicate that the nonlinear ensemble forecasting model performs better than the other models presented here.

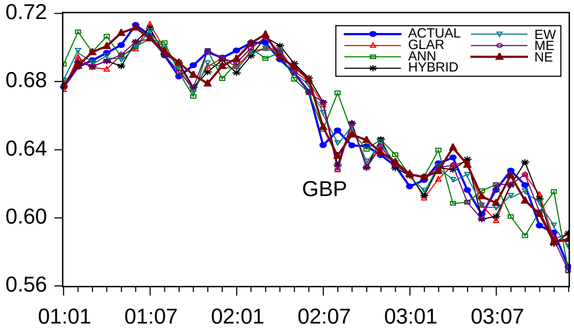

Fig. 3. A graphical comparison of GBP rate prediction results using different models (2001:01–2003:12).

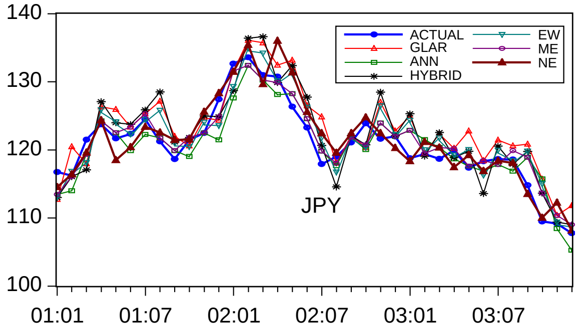

Fig. 4. A graphical comparison of JPY rate prediction results using different models (2001:01–2003:12).

Table 1
A comparison of NMSE between different methods for the three main currencies

Methods	Currencies					
	DEM		GBP		JPY	
	NMSE	Rank	NMSE	Rank	NMSE	Rank
GLAR	0.0453	5	0.0737	4	0.4207	6
ANN	0.0433	4	0.1031	6	0.1982	3
Hybrid	0.0459	6	0.0758	5	0.3854	5
EW	0.0319	3	0.0469	3	0.2573	4
ME	0.0143	1	0.0410	2	0.1505	2
NE	0.0156	2	0.0357	1	0.1391	1

GLAR: generalized linear auto-regression; ANN: artificial neural network; EW: equal weights; ME: minimum error; NE: nonlinear ensemble; NMSE: normalized mean square error.

Subsequently, the forecasting performance comparisons of various models for the three main currencies via NMSE, D_{stat} and return rate (R) are reported in Tables 1–3, respectively. Generally speaking, these tables provide comparisons of NMSE, D_{stat} and R between these different methods, indicating that the

Table 2
A comparison of D_{stat} between different methods for the three main currencies

Methods	Currencies					
	DEM		GBP		JPY	
	D_{stat} (%)	Rank	D_{stat} (%)	Rank	D_{stat} (%)	Rank
GLAR	58.33	6	58.33	6	47.22	6
ANN	63.89	4	69.44	3	63.89	4
Hybrid	61.11	5	61.11	5	66.67	3
EW	69.44	3	66.67	4	58.33	5
ME	80.56	2	77.78	2	72.22	2
NE	83.33	1	86.11	1	75.00	1

GLAR: generalized linear auto-regression; ANN: artificial neural network; EW: equal weights; ME: minimum error; NE: nonlinear ensemble; D_{stat}: directional statistics.

Table 3
A comparison of annual return rate (R) between different methods for the three main currencies

Methods	Currencies					
	DEM		GBP		JPY	
	R (%)	Rank	R (%)	Rank	R (%)	Rank
GLAR	5.89	6	5.14	6	−1.76	6
ANN	7.53	4	8.69	3	6.23	4
Hybrid	7.38	5	6.47	5	7.58	3
EW	8.65	3	7.53	4	5.68	5
ME	11.87	2	10.56	2	9.34	2
NE	15.54	1	14.85	1	16.49	1

GLAR: generalized linear auto-regression; ANN: artificial neural network; EW: equal weights; ME: minimum error; NE: nonlinear ensemble; R: annual return rate.

prediction performance of the nonlinear ensemble forecasting model is better than those of other models including these empirical analyses for the three main currencies.

Focusing on the NMSE indicator, the nonlinear ensemble model performs the best in all but the DEM case, followed by the linear combination model with ME method and EW method and other individual models. To summarize, the nonlinear ensemble model and the linear combination model with ME method outperform the other different models presented here in terms of NMSE. Interestingly, the NMSEs of the hybrid methodology are not better than those of the individual GLAR and ANN models. This result differs from that of Zhang [12]. This could be because the GLAR is different from ARIMA, or because the ANN architecture designs may differ.

However, the low NMSE does not necessarily mean that there is a high hit rate of forecasting direction for foreign exchange movement direction prediction. Thus, the D_{stat} comparison is necessary. Focusing on D_{stat} of Table 2, we find the nonlinear ensemble model also performs much better than the other models according to the rank; furthermore, from the business practitioners' point of view, D_{stat} is more important than NMSE because the former is an important decision criterion. With reference to Table 2, the differences between the different models are very significant. For example, for the GBP test case, the D_{stat} for the individual GLAR model is 58.33%, for the individual ANN model it is 69.44%, and for the single hybrid model D_{stat} is only 61.11%; while for the nonlinear ensemble forecasting model, D_{stat} reaches 86.11%. For the linear combination model with EW method and the hybrid method, the rank of forecasting accuracy is always in the middle for any of the test currencies. The main cause of this phenomenon is that the bad performance of the individual models has an important effect on the holistic forecast efficiency. Similarly, the individual ANN can model nonlinear time series such as exchange rate series well, and the D_{stat} rank is also in the middle for any of the test currencies. In the same way, we also notice that the D_{stat} ranking for GLAR is the last. The main reason is that the high noise, nonlinearity and complex factors are contained in foreign exchange series, and unfortunately the GLAR is a class of linear model.

In terms of the return or profit criterion, the empirical results show that the proposed nonlinear ensemble model could be applied to future forecasting. Compared with the other models presented in this paper, the nonlinear ensemble forecasting model performs the best. Likewise, we find that the rank of Table 3 is the similar to that of Table 2. It is not hard to understand such a rationale that right forecasting direction often leads to high return rates. As shown in Table 3, for the DEM test case the best return rate for the nonlinear ensemble model is 15.54%. Similarly, in the JPY test case, the annual return rate for GLAR is -1.76%, the rate for the ANN model is 6.23%, the rate for the hybrid model is 7.58%, and the rate for the linear combination model with ME method is also 9.34%; however the return rate for the nonlinear ensemble model reaches 16.49% per year.

From the experiments presented in this study we can draw the following conclusions. (i) The experimental results show that the nonlinear ensemble forecasting model is superior to the individual GLAR model, the individual ANN model, the single hybrid model as well as the linear combination models with EW and ME methods for the test cases of three main currencies in terms of the measurement of directional change statistics (D_{stat}) and annual return rate (R), as can be seen from Tables 2 and 3. Likewise, the nonlinear ensemble model and the linear combination model with ME method also outperform other models in terms of goodness-of-fit or NMSE, as can be seen from Figs. 2–4 and Table 1. (ii) The nonlinear ensemble forecasts are able to improve forecasting accuracy significantly—in other words, the performance of the nonlinear ensemble forecasting model is better than those of other forecasting models in terms of NMSE, D_{stat} and R. This leads to the third conclusion. (iii) The nonlinear ensemble model can be used as an alternative tool for exchange rate forecasting to obtain greater forecasting accuracy and improve the prediction quality further in view of empirical results.

4. Conclusions

This study proposes using a nonlinear ensemble forecasting model that combines the time series GLAR model, the ANN model and the hybrid model to predict foreign exchange rates. In terms of the empirical results, we find that across different forecasting models for the test cases of three main currencies—German marks, British pounds and Japanese yen—on the basis of different criteria, the nonlinear ensemble model performs the best. In the nonlinear ensemble model test cases, the NMSE is the lowest and the D_{stat} and R is the highest, indicating that the nonlinear ensemble forecasting model can be used as a viable alternative solution for exchange rate prediction for financial managers and business practitioners.

References

[1] De Matos G. Neural networks for forecasting exchange rate. M. Sc. thesis, The University of Manitoba, Canada, 1994.
[2] Kuan CM, Liu T. Forecasting exchange rates using feedforward and recurrent neural networks. Journal of Applied Econometrics 1995;10:347–64.
[3] Hsu W, Hsu LS, Tenorio MF. A neural network procedure for selecting predictive indicators in currency trading. In: Refenes AN., editor. Neural networks in the capital markets. New York: Wiley; 1995. pp. 245–57.
[4] Tenti P. Forecasting foreign exchange rates using recurrent neural networks. Applied Artificial Intelligence 1996;10:567–81.
[5] El Shazly MR, El Shazly HE. Forecasting currency prices using a genetically evolved neural network architecture. International Review of Financial Analysis 1999;8:67–82.
[6] Leung MT, Chen AS, Daouk H. Forecasting exchange rates using general regression neural networks. Computers & Operations Research 2000;27:1093–110.
[7] Zhang GP, Berardi VL. Time series forecasting with neural network ensembles: an application for exchange rate prediction. Journal of the Operational Research Society 2001;52:652–64.
[8] Wedding II DK, Cios KJ. Time series forecasting by combining RBF networks, certainty factors, and the Box-Jenkins model. Neurocomputing 1996;10:149–68.
[9] Luxhoj JT, Riis JO, Stensballe B. A hybrid econometric-neural network modeling approach for sales forecasting. International Journal of Production Economics 1996;43:175–92.
[10] Voort MVD, Dougherty M, Watson S. Combining Kohonen maps with ARIMA time series models to forecast traffic flow. Transportation Research Part C: Emerging Technologies 1996;4:307–18.
[11] Tseng FM, Yu HC, Tzeng GH. Combining neural network model with seasonal time series ARIMA model. Technological Forecasting and Social Change 2002;69:71–87.
[12] Zhang GP. Time series forecasting using a hybrid ARIMA and neural network model. Neurocomputing 2003;50:159–75.
[13] Reid DJ. Combining three estimates of gross domestic product. Economica 1968;35:431–44.
[14] Bates JM, Granger CWJ. The combination of forecasts. Operations Research Quarterly 1969;20:451–68.
[15] Makridakis S, Anderson A, Carbone R, Fildes R, Hibdon M, Lewandowski R, Newton J, Parzen E, Winkler R. The accuracy of extrapolation (time series) methods: results of a forecasting competition. Journal of Forecasting 1982;1:111–53.
[16] Pelikan E, De Groot C, Wurtz D. Power consumption in West-Bohemia: improved forecasts with decorrelating connectionist networks. Neural Network World 1992;2:701–12.
[17] Ginzburg I, Horn D. Combined neural networks for time series analysis. Neural Information Processing Systems 1994;6:224–31.
[18] Clemen R. Combining forecasts: a review and annotated bibliography with discussion. International Journal of Forecasting 1989;5:559–608.
[19] Pickett J, editor. The American heritage dictionary, 4th ed. Boston: Houghton Mifflin; 2000.
[20] Shephard N. Generalized linear autoregressions. Economics working paper 8, Nuffield College, Oxford, 1995.
[21] Gujarati DN. Basic econometrics, 3rd ed. New York: McGraw-Hill, Inc.; 1995. p. 741–46.
[22] Box GEP, Jenkins G. Time series analysis: forecasting and control. San Francisco, CA: Holden-Day; 1970.
[23] Firth D. Generalized linear models. In: Hinkley DV, Reid N, Snell EJ., editors. Statistical theory and modelling. London: Chapman & Hall; 1990. pp. 55–82.
[24] McCullagh P, Nelder JA. Generalized linear models. 2nd ed., London: Chapman & Hall; 1989.
[25] Tang Z, Almeida C, Fishwick PA. Time series forecasting using neural networks vs. Box–Jenkins methodology. Simulation 1991;57:303–10.
[26] Tang Z, Fishwick PA. Feedforward neural nets as models for time series forecasting. ORSA Journal on Computing 1993;5:374–85.
[27] Zhang GP, Patuwo EB, Hu MY. A simulation study of artificial neural networks for nonlinear time-series forecasting. Computers & Operations Research 2001;28:381–96.

[28] Hwang HB. Insights into neural-network forecasting time series corresponding to ARMA (p, q) structures. Omega 2001;29:273–89.
[29] Denton JW. How good are neural networks for causal forecasting?. Journal of Business Forecasting 1995;14:17–20.
[30] Markham IS, Rakes TR. The effect of sample size and variability of data on the comparative performance of artificial neural networks and regression. Computers & Operations Research 1998;25:251–63.
[31] Chen AS, Leung MT. Regression neural network for error correction in foreign exchange rate forecasting and trading. Computers & Operations Research 2004;31(7):1049–68.
[32] Batchelor R, Dua P. Forecaster diversity and the benefits of combining forecasts. Management Science 1995;41:68–75.
[33] Kang BH. Unstable weights in the combination of forecasts. Management Science 1986;32:683–95.
[34] Shi SM, Xu LD, Liu B. Improving the accuracy of nonlinear combined forecasting using neural networks. Expert Systems with Application 1999;16:49–54.
[35] Yang GY, Tang XW, Ma YK. The discussions of many problems about combined forecasting with non-negative weights. Journal of XiDian University 1996;25(2):210–5.
[36] Jolliffe IT. Principal component analysis. Berlin: Springer; 1986.
[37] Karhunen J, Joutsensalo J. Generalizations of principal component analysis, optimization problems, and neural networks. Neural Networks 1995;8(4):549–62.
[38] Weigend AS, Gershenfeld NA, editors. Time series prediction: forecasting the future and understanding the past. Proceedings of the NATO advanced research workshop on comparative time series analysis, Santa Fe, NM, May 14–17, 1992. Reading, MA: Addison-Wesley; 1994.
[39] Yao JT, Tan CL. A case study on using neural networks to perform technical forecasting of forex. Neurocomputing 2000;34:79–98.