Journal of Industrial and Systems Engineering
Vol. 4, No. 4, pp 265-285
Winter 2011

Which Methodology is Better for Combining Linear and Nonlinear Models for Time Series Forecasting?

Mehdi Khashei[1*], Mehdi Bijari[2]

[1,2]Department of Industrial and Systems Engineering, Isfahan University of Technology, Isfahan, Iran

[1] khashei@in.iut.ac.ir

ABSTRACT

Both theoretical and empirical findings have suggested that combining different models can be an effective way to improve the predictive performance of each individual model. It is especially occurred when the models in the ensemble are quite different. Hybrid techniques that decompose a time series into its linear and nonlinear components are one of the most important kinds of the hybrid models for time series forecasting. Several researches in the literature have been shown that these models can outperform single models. In this paper, the predictive capabilities of three different models in which the autoregressive integrated moving average (ARIMA) as linear model is combined to the multilayer perceptron (MLP) as nonlinear model, are compared together for time series forecasting. These models are including the Zhang's hybrid ANNs/ARIMA, artificial neural network (p,d,q), and generalized hybrid ANNs/ARIMA models. The empirical results with three well-known real data sets indicate that all of these methodologies can be effective ways to improve forecasting accuracy achieved by either of components used separately. However, the generalized hybrid ANNs/ARIMA model is more accurate and performs significantly better than other aforementioned models.

Keywords: Artificial Neural Networks (ANNs), Auto-Regressive Integrated Moving Average (ARIMA), Time series forecasting, Hybrid linear/nonlinear models.

1. INTRODUCTION

Applying quantitative models for forecasting and assisting investment decision making has become more indispensable in many areas. Time series forecasting is one of the most important types of quantitative models in which past observations of the same variable are collected and analyzed to develop a model describing the underlying relationship (Aryal & Yao-Wu, 2003). This modeling approach is particularly useful when little knowledge is available on the underlying data generating process or when there is no satisfactory explanatory model that relates the prediction variable to other explanatory variables (Zhang, 2003). Forecasting procedures include different techniques and models. Moving averages techniques, random walks and trend models, exponential smoothing, state space modeling, multivariate methods, vector autoregressive models, cointegrated and causal

∗ Corresponding Author

ISSN: 1735-8272,

models, methods based on neural, fuzzy networks or data mining and rule-based techniques are typical models used in time series forecasting (Ragulskis & Lukoseviciute, 2009).

Auto-regressive integrated moving average (ARIMA) models are one of the most important and widely used linear time series models. The popularity of the ARIMA model is due to its statistical properties as well as the well-known Box–Jenkins methodology (Box & Jenkins, 1976) in the model building process. In addition, various exponential smoothing models can be implemented by ARIMA models. Although ARIMA models are quite flexible in that they can represent several different types of time series and also have the advantages of accurate forecasting over a short period of time and ease of implementation, their major limitation is the pre-assumed linear form of the model. ARIMA models assume that future values of a time series have a linear relationship with current and past values as well as with white noise, so approximations by ARIMA models may not be adequate for complex nonlinear real-world problems. However, many researchers have argued that real world systems are often nonlinear (Zhang *et al.*, 1998). These evidences have encouraged academic researchers and business practitioners in order to develop more predictable forecasting models than linear models (Khashei & Bijari, 2011).

Several classes of parametric and nonparametric nonlinear models have been proposed in the literature in order to overcome the restriction of the linear models and to account nonlinear patterns observed in real problems. Among parametric models, the bilinear model (Granger & Anderson, 1978), the threshold autoregressive (TAR) model (Tong & Lim, 1980), the autoregressive conditional heteroscedastic (ARCH) model (Engle, 1982) and generalized autoregressive conditional heteroscedastic (GARCH) model (Bollerslev, 1986), chaotic dynamics (Hsieh, 1991), and self-exciting threshold autoregressive (Chappel *et al.*, 1996) receive the most attention. While these models may be good for a particular situation, they perform poorly for other applications. The reason is that the above-mentioned models are developed for specific nonlinear patterns and are not capable of modeling other types of nonlinearity in time series (Khashei & Bijari, 2011).

A number of nonparametric forecasting models such as multivariate nearest-neighbor methods have also been proposed for time series forecasting. However, the results of some researchers investigated in time series forecasting suggest that these nonparametric models cannot significantly improve forecasts accuracy upon the other time series models (Mizrach, 1992). Artificial neural networks (ANNs) are another type of nonparametric nonlinear models, which have been proposed and examined for time series forecasting. Given the advantages of neural networks (Panda & Narasimhan, 2007), it is not surprising that this methodology has attracted overwhelming attention in time series forecasting. Artificial neural networks have been found to be a viable contender to various traditional time series models (Chen *et al.*, 2005; Giordano *et al.*, 2007; Jain & Kumar, 2007). Lapedes and Farber (1987) report the first attempt to model nonlinear time series with artificial neural networks. De Groot and Wurtz (1991) present a detailed analysis of univariate time series forecasting using feedforward neural networks for two benchmark nonlinear time series. Chakraborty *et al.* (1992) conduct an empirical study on multivariate time series forecasting with artificial neural networks. Poli and Jones (1994) propose a stochastic neural network model based on Kalman filter for nonlinear time series prediction. Cottrell *et al.* (1995) address the issue of network structure for forecasting real world time series. Berardi and Zhang (2003) investigate the bias and variance issue in the time series forecasting context. In addition, several large forecasting competitions (Balkin & Ord, 2000; Weigend & Gershenfeld, 1993) suggest that neural networks can be a very useful addition to the time series forecasting toolbox. Santos *et al.* (2007) investigate the hypothesis that the nonlinear mathematical models of multilayer perceptron and the radial basis function neural networks are able to provide a more accurate out-of-sample forecast than the traditional linear models. Their results indicate that ANNs perform better than their linear models (Khashei & Bijari, 2011).

Although artificial neural networks have the advantages of accurate forecasting, their performance in some specific situations is inconsistent. In the literature, several papers are devoted to comparing ANNs with the traditional methods. Despite the numerous studies, which have shown ANNs are significantly better than the conventional linear models and their forecast considerably and consistently more accurately, some other studies have reported inconsistent results. Foster *et al.* (1992) find that ANNs are significantly inferior to linear regression and a simple average of exponential smoothing methods. Brace *et al.* (1991) also find that the performance of ANNs is not as good as many other statistical methods commonly used in the load forecasting. Denton (1995) with generated data for several different experimental conditions shows that under ideal conditions, with all regression assumptions, there is little difference in the predictability between ANNs and linear regression, and only under less ideal conditions such as outliers, multicollinearity, and model misspecification, ANNs perform better. Hann and Steurer (1996) make comparisons between the neural networks and the linear model in exchange rate forecasting. They report that if monthly data are used, neural networks do not show much improvement over linear models. Taskaya and Casey (2005) compare the performance of linear models with neural networks. Their results show that linear autoregressive models can outperform neural networks in some cases (Khashei & Bijari, 2011).

Most other researchers also make comparisons between ANNs and the corresponding traditional methods in their particular applications. Fishwick (1989) reports that the performance of ANNs is worse than that of the simple linear regression. Tang *et al.* (1991), and Tang and Fishwick (1993) try to answer the question: under what conditions ANN forecasters can perform better than the linear time series forecasting methods such as Box- Jenkins models. Some researchers believe that in some specific situations where ANNs perform worse than linear statistical models, the reason may simply be that the data is linear without much disturbance, therefore; cannot be expected that ANNs to do better than linear models for linear relationships (Zhang *et al.*, 1998). However, for any reason, using ANNs to model linear problems have yielded mixed results and hence; it is not wise to apply ANNs blindly to any type of data.

One of the major developments in neural networks over the last decade is the model combining or ensemble modeling. The basic idea of this multi-model approach is the use of each component model's unique capability to better capture different patterns in the data. Both theoretical and empirical findings have suggested that combining different models can be an effective way to improve the predictive performance of each individual model, especially when the models in the ensemble are quite different (Zhang, 2007). In addition, since it is difficult to completely know the characteristics of the data in a real problem, hybrid methodology that has both linear and nonlinear modeling capabilities can be a good strategy for practical use. Although a majority of the neural ensemble literature is focused on pattern classification problems, a number of combining schemes have been proposed for time series forecasting problems (Zou *et al.*, 2007).

The literature of hybrid models for time series forecasting has dramatically expanded since the early work of Reid (1968), and Bates and Granger (1969). Clemen (1989) provided a comprehensive review and annotated bibliography in this area. Wedding and Cios (1996) described a combining methodology using radial basis function networks (RBF) and the Box–Jenkins ARIMA models. Tsaih *et al.* (1998) presented a hybrid artificial intelligence integrating the rule-based systems technique and the neural networks technique to predict accurately the direction of daily price changes in S&P 500 stock index futures. Pelikan *et al.* (1992), and Ginzburg and Horn (1994) proposed to combine several feedforward neural networks in order to improve time series forecasting accuracy. Luxhoj *et al.* (1996) presented a hybrid econometric and ANN approach for sales forecasting. Goh *et al.* (2003) used an ensemble of boosted Elman networks for predicting drug dissolution profiles. Voort *et al.* introduced a hybrid method called KARIMA using a Kohonen self-organizing map and autoregressive integrated moving average method for short-term prediction (1996). Medeiros and Veiga (1989) consider a hybrid time series forecasting system with neural networks used to control the time-varying parameters of a smooth transition autoregressive model. Armano *et al.* (2005) presented a new hybrid approach that integrated artificial neural network with genetic algorithms (GAs) to stock market forecast.

In recent years, more hybrid forecasting models have been proposed, using autoregressive integrated moving average and artificial neural networks and applied to time series forecasting with good prediction performance. Pai and Lin (2005) proposed a hybrid methodology to exploit the unique strength of ARIMA models and Support Vector Machines (SVMs) for stock prices forecasting. Chen and Wang (2007) constructed a combination model incorporating seasonal autoregressive integrated moving average (SARIMA) model and SVMs for seasonal time series forecasting. Zhou and Hu (2008) proposed a hybrid modeling and forecasting approach based on Grey and Box–Jenkins autoregressive moving average (ARMA) models. Khashei *et al.* (2009) presented a hybrid ARIMA and artificial intelligence approaches to financial markets prediction. Yu *et al.* (2005) proposed a novel nonlinear ensemble forecasting model integrating generalized linear auto regression (GLAR) with artificial neural networks in order to obtain accurate prediction in foreign exchange market. Kim and Shin (2007) investigated the effectiveness of a hybrid approach based on the artificial neural networks for time series properties, such as the adaptive time delay neural networks (ATNNs) and the time delay neural networks (TDNNs), with the genetic algorithms in detecting temporal patterns for stock market prediction tasks. Tseng *et al.* (2002) proposed using a hybrid model called SARIMABP that combines the seasonal autoregressive integrated moving average (SARIMA) model and the back-propagation neural network model to predict seasonal time series data. Khashei *et al.* (2008) based on the basic concepts of artificial neural networks, proposed a new hybrid model in order to overcome the data limitation of neural networks and yield more accurate forecasting model, especially in incomplete data situations.

Hybrid techniques that decompose a time series into its linear and nonlinear form are one of the most popular hybrid models, which have recently been shown to be successful for single models. The linear ARIMA and the nonlinear multilayer perceptrons are jointly used in these hybrid models in order to capture different forms of relationship in the time series data. The motivation of these hybrid models come from the following perspectives. First, it is often difficult in practice to determine whether a time series under study is generated from a linear or nonlinear underlying process; thus, the problem of model selection can be eased by combining linear ARIMA and nonlinear ANN models. Second, real-world time series are rarely pure linear or nonlinear and often contain both linear and nonlinear patterns, which neither ARIMA nor ANN models alone can be adequate for modeling in such cases; hence the problem of modeling the combined linear and nonlinear autocorrelation structures in time series can be solved by combining linear ARIMA and nonlinear ANN models. Third, it is almost universally agreed in the forecasting literature that no single model is the best in every situation, due to the fact that a real-world problem is often complex in nature and any single model may not be able to capture different patterns equally well. Therefore, the chance in order to capture different patterns in the data can be increased by combining different models (Zhang, 2003).

In this paper, three different methodologies that have been proposed in order to combine the autoregressive integrated moving average (ARIMA) as linear model and multilayer perceptron

(MLP) as nonlinear model are presented. Moreover, the predictive capabilities of the constructed models based on these methodologies for time series forecasting— Zhang's hybrid ANNs/ARIMA (Zhang, 2003), artificial neural network (p,d,q) (Khashei & Bijari, 2010), and generalized hybrid ANNs/ARIMA—are compared together and also their components, using three well-known real data sets. The data sets are including the Wolf's sunspot data, the Canadian lynx data, and the British pound against the United States dollar exchange rate data. The rest of the paper is organized as follows. In the next section, the basic concepts and modeling approaches of the autoregressive integrated moving average (ARIMA) models, artificial neural networks (ANNs), and the above-mentioned hybrid models are briefly introduced. Description of used data sets is presented in section 3. Empirical results of above-mentioned hybrid models for time series forecasting from three real data sets are reported in Section 4. Section 5 contains the concluding remarks.

2. THE AUTOREGRESSVE INTEGRATED MOVING AVERAGE, ARTIFICIAL NEURAL NETWORKS, AND HYBRID ANNs/ARIMA MODELS

In this section, the basic concepts and modeling approaches of the autoregressive integrated moving average (ARIMA), artificial neural networks (ANNs), and hybrid artificial neural networks and autoregressive integrated moving average models for time series forecasting are briefly reviewed.

2.1. The autoregressive integrated moving average (ARIMA) models

For more than half a century, autoregressive integrated moving average (ARIMA) models have dominated many areas of time series forecasting. In an autoregressive integrated moving average (*p*,*d*,*q*) model, the future value of a variable is assumed to be a linear function of several past observations and random errors. That is, the underlying process that generates the time series with the mean μ has the form (Khashei & Bijari, 2010).

$$\phi(B)\nabla^d (y_t - \mu) = \theta(B)a_t \tag{1}$$

where, y_t and a_t are the actual value and random error at time period *t*, respectively; $\phi(B) = 1 - \sum_{i=1}^{p} \varphi_i B^i$, $\theta(B) = 1 - \sum_{j=1}^{q} \theta_j B^j$ are polynomials in *B* of degree *p* and *q*, $\phi_i \quad (i = 1,2,\dots,p)$ and $\theta_j \quad (j = 1,2,\dots,q)$ are model parameters, $\nabla = (1 - B)$, *B* is the backward shift operator, *p* and *q* are integers and often referred to as orders of the model, and *d* is an integer and often referred to as order of differencing. Random errors, a_t, are assumed to be independently and identically distributed with a mean of zero and a constant variance of σ^2 (Khashei & Bijari, 2010).

Based on the earlier work of Yule (1926) and Wold (1938), Box and Jenkins (1976) developed a practical approach to building ARIMA models, which has the fundamental impact on the time series analysis and forecasting applications. The Box–Jenkins methodology includes three iterative steps of model identification, parameter estimation, and diagnostic checking. The basic idea of model identification is that if a time series is generated from an autoregressive integrated moving average process, it should have some theoretical autocorrelation properties. By matching the empirical autocorrelation patterns with the theoretical ones, it is often possible to identify one or several potential models for the given time series. Box and Jenkins (1976) proposed to use the autocorrelation function (ACF) and the partial autocorrelation function (PACF) of the sample data as the basic tools to identify the order of the autoregressive integrated moving average model. Some other order selection methods have been proposed based on validity criteria, the informationtheoretic approaches such as the Akaike's information criterion (AIC) (Shibata, 1976) and the minimum description length (MDL) (Jones, 1975; Hurvich & Tsai, 1989; Ljung, 1987). In addition, in recent years different approaches based on intelligent paradigms, such as neural networks (Hwang, 2001), genetic algorithms (Minerva & Poli, 2001; Ong *et al.*, 2005) or fuzzy system (Haseyama & Kitajima, 2001) have been proposed to improve the accuracy of order selection of ARIMA models (Khashei & Bijari, 2010).

In the identification step, data transformation is often required to make the time series stationary. Stationarity is a necessary condition in building an autoregressive integrated moving average model used for forecasting. A stationary time series is characterized by statistical characteristics such as the mean and the autocorrelation structure being constant over time. When the observed time series presents trend and heteroscedasticity, differencing and power transformation are applied to the data to remove the trend and to stabilize the variance before an autoregressive integrated moving average model can be fitted. Once a tentative model is identified, estimation of the model parameters is straightforward. The parameters are estimated such that an overall measure of errors is minimized. This can be accomplished using a nonlinear optimization procedure. The last step in model building is the diagnostic checking of model adequacy. This is basically to check if the model assumptions about the errors, a_t , are satisfied (Khashei & Bijari, 2010).

Several diagnostic statistics and plots of the residuals can be used to examine the goodness of fit of the tentatively entertained model to the historical data. If the model is not adequate, a new tentative model should be identified, which will again be followed by the steps of parameter estimation and model verification. Diagnostic information may help suggest alternative model(s). This three-step model building process is typically repeated several times until a satisfactory model is finally selected. The final selected model can then be used for prediction purposes (Khashei & Bijari, 2010).

2.2. The artificial neural networks (ANNs) models

Recently, computational intelligence systems and among them artificial neural networks (ANNs), which in fact are model free dynamics, has been used widely for approximation functions and forecasting. One of the most significant advantages of the ANN models over other classes of nonlinear models is that ANNs are universal approximators that can approximate a large class of functions with a high degree of accuracy (Zhang *et al.*, 1998). Their power comes from the parallel processing of the information from the data. No prior assumption of the model form is required in the model building process. Instead, the network model is largely determined by the characteristics of the data. Single hidden layer feed forward network is the most widely used model form for time series modeling and forecasting. The model is characterized by a network of three layers of simple processing units connected by acyclic links (Figure 1). The relationship between the output (y_t) and the inputs ($y_{t-1},\ldots,y_{t-p}$) has the following mathematical representation (Khashei & Bijari, 2010).

$$y_t = w_0 + \sum_{j=1}^{q} w_j \cdot g\left(w_{0,j} + \sum_{i=1}^{p} w_{i,j} \cdot y_{t-i} \right) + \varepsilon_t , \qquad (2)$$

where, $w_{i,j}\,(i=0,1,2,\ldots,p,\;\; j=1,2,\ldots,q)$ and $w_j\,(j=0,1,2,\ldots,q)$ are model parameters often called connection weights; p is the number of input nodes; and q is the number of hidden nodes. Activation functions can take several forms. The type of activation function is indicated by the situation of the neuron within the network. In the majority of cases input layer neurons do not have an activation function, as their role is to transfer the inputs to the hidden layer. The most widely used activation function for the output layer is the linear function as non-linear activation function may introduce distortion to the predicated output. The logistic function is often used as the hidden layer transfer function that are shown in Eq. 3. Other activation functions can also be used such as linear and quadratic, each with a variety of modeling applications (Khashei & Bijari, 2010).

$$Sig(x) = \frac{1}{1+exp(-x)}. \tag{3}$$

Hence, the ANN model of (2), in fact, performs a nonlinear functional mapping from past observations to the future value y_t , i.e.,

$$y_t = f(y_{t-1},..., y_{t-p}, w) + \varepsilon_t , \tag{4}$$

where, w is a vector of all parameters and $f(.)$ is a function determined by the network structure and connection weights. Thus, the neural network is equivalent to a nonlinear autoregressive model. The simple network given by (2) is surprisingly powerful in that it is able to approximate the arbitrary function as the number of hidden nodes when q is sufficiently large. In practice, simple network structure that has a small number of hidden nodes often works well in out-of-sample forecasting. This may be due to the overfitting effect typically found in the neural network modeling process. An overfitted model has a good fit to the sample used for model building but has poor generalizability to data out of the sample (Khashei & Bijari, 2010).

The choice of q is data-dependent and there is no systematic rule in deciding this parameter. In addition to choosing an appropriate number of hidden nodes, another important task of ANN modeling of a time series is the selection of the number of lagged observations, p, and the dimension of the input vector. This is perhaps the most important parameter to be estimated in an ANN model because it plays a major role in determining the (nonlinear) autocorrelation structure of the time series (Khashei & Bijari, 2010).

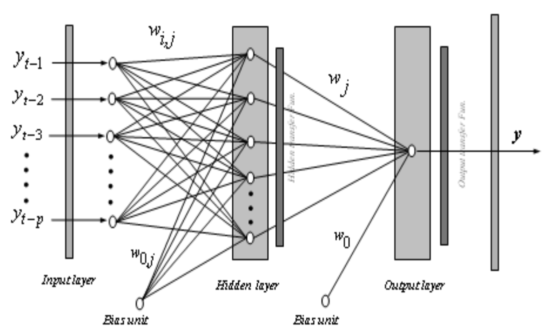

Figure 1 Architecture of a neural network in the general form (***N*** $^{(p-q-1)}$.)

Although many different approaches exist in order to find the optimal architecture of a neural network, these methods are usually quite complex in nature and are difficult to implement. Furthermore, none of these methods can guarantee the optimal solution for all real forecasting problems. To date, there is no simple clear-cut method for determination of these parameters and the usual procedure is to test numerous networks with varying numbers of input and hidden units, estimate generalization error for each, and select the network with the lowest generalization error (Khashei & Bijari, 2010).

2.3. Hybrid artificial neural networks and autoregressive integrated moving average models

Both ARIMA and ANN models have achieved successes in their own linear or nonlinear domains. However, none of them is a universal model that is suitable for all circumstances. The approximation of ARIMA models to complex nonlinear problems may not be adequate. On the other hand, using ANNs to model linear problems have yielded mixed results. Hence, it is not wise to apply ANNs blindly to any type of data. Since it is difficult to completely know the characteristics of the data in a real problem, hybrid methodology that has both linear and nonlinear modeling capabilities can be a good strategy for practical use. By combining different models, different aspects of the underlying patterns may be captured (Zhang, 2003).

2.3.1. Zhang's hybrid ANNs/ARIMA model

Some researchers in hybrid linear and nonlinear models believe that it may be reasonable to consider a time series to be composed of a linear autocorrelation structure and a nonlinear component (Zhang, 2003). That is,

$$y_t = N_t + L_t \tag{5}$$

where L_t denotes the linear component and N_t denotes the nonlinear component. These two components have to be estimated from the data. First, we let ARIMA to model the linear component, and then the residuals from the linear model will contain only the nonlinear relationship (Zhang, 2003). Let e_t denote the residual at time *t* from the linear model, then

$$e_t = y_t - \hat{L}_t \tag{6}$$

where $\hat{L}_t$ is the forecast value for time *t* from the estimated relationship (1). By modeling residuals using ANNs, nonlinear relationships can be discovered (Zhang, 2003). With *n* input nodes, the ANN model for the residuals will be:

$$e_t = f(e_{t-1}, \ldots, e_{t-n}) + \varepsilon_t \tag{7}$$

where *f* is a nonlinear function determined by the neural network and e_t is the random error. Note that if the model *f* is not an appropriate one, the error term is not necessarily random (Zhang, 2003). Therefore, the correct model identification is critical. Denote the forecast from (7) as $\hat{N}_t$, the combined forecast will be

$$\hat{y}_t = \hat{L}_t + \hat{N}_t \tag{8}$$

The hybrid model exploits the unique feature and strength of ARIMA model as well as ANN model in determining different patterns. Thus, it could be advantageous to model linear and nonlinear patterns separately by using different models and then combine the forecasts to improve the overall modeling and forecasting performance (Zhang, 2003).

2.3.1.1. Advantages and disadvantages

The performance of the Zhang's hybrid model is often satisfactory than each component model used in isolation (Taskaya & Ahmad, 2005). In addition, it can be generally guaranteed that the performance of the Zhang's hybrid model will not be worse than ARIMA model. However, despite the all advantages mentioned for Zhang's hybrid model, it has some assumptions that will degenerate its performance if the opposite situations occur. These assumptions are as follows:

1- This model supposes that the linear and nonlinear patterns of a time series can be separately modeled by different models and then the forecasts can be combined together and this may degrade performance, if it is not true.

2- This model supposes that the relationship between the linear and nonlinear components is additive and this may underestimate the relationship between the components and degrade performance, if there is not any additive association between the linear and nonlinear elements and the relationship is different (Taskaya & Casey, 2005).

3- This model supposes that the residuals from the linear model will contain only the nonlinear relationship. However, one may not guarantee that the residuals of the linear component may comprise valid nonlinear patterns (Taskaya & Casey, 2005).

In addition, as mentioned previously, it cannot be generally guaranteed that the performance of the Zhang's hybrid model will not be worse than ANN model.

2.3.2. An artificial neural network (p, d, q) model

Although traditional hybrid linear and nonlinear models such as Zhang's hybrid model have recently been shown to be successful for single models, perhaps the danger in using these hybrid models is that there are some assumptions considered in constructing process of these hybrid models that will degenerate their performance if the opposite situations occur. Therefore, they may be inadequate in some specific situations. For example, in these models are assumed that the existing linear and nonlinear patterns in a time series can be separately modeled or the residuals from the linear model contain only the nonlinear relationship or the relationship between the linear and nonlinear components is additive. Therefore, these assumptions may underestimate the relationship between the components and degrade performance, if the opposite situation occurs, for example, if the existing linear and nonlinear patterns in a time series cannot be separately modeled or the residuals of the linear component don't comprise valid nonlinear patterns or is not any additive association between the linear and nonlinear elements and the relationship is different (for example multiplicative). In addition, as mentioned previously, it cannot be generally guaranteed that the performance of these hybrid models will be better than both component models (Khashei & Bijari, 2010).

Artificial neural network (p,d,q) model is proposed in order to overcome the above-mentioned limitations of the traditional hybrid linear and nonlinear models such as Zhang's hybrid model. This model also is a hybrid linear and nonlinear model that combines an autoregressive integrated moving average (ARIMA) as linear model with a multilayer perceptron as nonlinear model using a new methodology in order to yield more accurate results. In the artificial neural network (p,d,q) model such as in the Box–Jenkins methodology in linear modeling, the future value of a time series is considered as nonlinear function of several past observations and random errors as follows (Khashei & Bijari, 2010).

$$y_t = f\left[(z_{t-1}, z_{t-2}, \ldots z_{t-m}), (e_{t-1}, e_{t-2}, \ldots e_{t-n})\right] \tag{9}$$

where f is a nonlinear function determined by the neural network, $z_t = (1-B)^d (y_t - \mu)$, e_t is the residual of the ARIMA model at time t and m and n are integers. So, in the first stage, an autoregressive integrated moving average model is used in order to generate the residuals (e_t).

In second stage, a neural network is used in order to model the nonlinear and linear relationships existing in residuals and original observations. Thus,

$$z_t = w_0 + \sum_{j=1}^{Q} w_j \cdot g(w_{0,j} + \sum_{i=1}^{p} w_{i,j} \cdot z_{t-i} + \sum_{i=p+1}^{p+q} w_{i,j} \cdot e_{t+p-i}) + \varepsilon_t, \tag{10}$$

where, $w_{i,j} (i = 0,1,2,..., p+q,\ j = 1,2,...,Q)$ and $w_j (j = 0,1,2,...,Q)$ are connection weights; p, q, Q are integers, which are determined in design process of final neural network (Khashei & Bijari, 2010). It must be noted that any set of above–mentioned variables $\{ e_i (i = t-1,...,t-n)\}$ or $\{ z_i (i = t-1,...,t-m)\}$ may be deleted in design process of final neural network. This maybe related to the underlying data generating process and the existing linear and nonlinear structures in data. For example, if data only consist of pure nonlinear structure, then the residuals will only contain the nonlinear relationship. For the reason that autoregressive integrated moving average is a linear model and does not able to model nonlinear relationship; therefore, the set of residuals $\{ e_i (i = t-1,...,t-n)\}$ variables maybe deleted against other of those variables (Khashei & Bijari, 2010).

2.3.2.1. Advantages and disadvantages

It can be seen that in the artificial neural network (p,d,q) model in contrast of the traditional hybrid models such as Zhang's hybrid model, no assumption is required in constructing process. In the artificial neural network (p,d,q) model is not needed to be assumed that the existing linear and nonlinear patterns in a time series can be separately modeled and they modeled simultaneously; or the residuals from the linear model only contain the nonlinear relationship. In addition, in this model, no prior assumption is considered for the relationship between the linear and nonlinear components and it will be generally estimated as function by neural network.

In additional, it can be generally guaranteed that the performance of the artificial neural network (p,d,q) model will not be worse than either of the components —autoregressive integrated moving average (ARIMA) and artificial neural networks (ANNs)— used separately. However, despite the all advantages mentioned for the artificial neural network (p,d,q) model, it cannot be generally guaranteed that the performance of this model will be better than the Zhang's hybrid model.

2.3.3. The generalized hybrid ANNs/ARIMA model

In order to yield a more general and more accurate hybrid linear and nonlinear model than the artificial neural network (p,d,q) model, generalized hybrid ANNs/ARIMA model has been proposed. The generalized hybrid ANNs/ARIMA model such as the artificial neural network (p,d,q) has no above-mentioned assumption of the traditional hybrid ARIMA and ANNs models. In this model, a time series is also considered as function of a linear and a nonlinear component. Thus,

$$y_t = f(L_t, N_t), \tag{11}$$

where L_t denotes the linear component and N_t denotes the nonlinear component. In the first stage, the main aim is linear modeling; therefore, an autoregressive integrated moving average (ARIMA) model is used to model the linear component. The residuals from the first stage will contain the nonlinear relationship that linear model dose not able to model it, and maybe linear relationship (Taskaya & Ahmad, 2005). Thus the L_t will be as follows.

$$L_t = \left[\sum_{i=1}^{p} \phi_i \, z_{t-i} - \sum_{j=1}^{q} \theta_j \, \varepsilon_{t-j} \right] + e_t = \hat{L}_t + e_t , \tag{12}$$

where $\hat{L}_t$ is the forecast value for time t from the estimated relationship (1), $z_t = (1-B)^d (y_t - \mu)$, and e_t is the residual at time t from the linear model. The forecasted values and residuals of linear modeling are the results of first stage that are used in next stage. In addition, the linear patterns are magnified by ARIMA model in order to apply in second stage.

In second stage, the main aim is nonlinear modeling; therefore, a multilayer perceptron is used in order to simultaneously model the nonlinear and probable linear relationships that may be remained in residuals of linear modeling and also the nonlinear and linear relationships in the original data. Thus,

$$N^1{}_t = f^1(e_{t-1},\ldots,e_{t-n}), \tag{13}$$

$$N^2{}_t = f^2(z_{t-1},\ldots,z_{t-m}), \tag{14}$$

$$N_t = f\left(N_t^{\,1}, N_t^{\,2}\right), \tag{15}$$

where f^1, f^2, and f are the nonlinear functions determined by the neural network. n and m are integers and are often referred to as orders of the model. Thus, the combined forecast will be as follows:

$$y_t = f(N_t^{\,1}, \hat{L}_t, N_t^{\,2}) = f(e_{t-1},\ldots,e_{t-n_1}, \hat{L}_t, z_{t-1},\ldots,z_{t-m_1}) \tag{16}$$

where f are the nonlinear functions determined by the neural network. $n_1 \le n$ and $m_1 \le m$ are integers determined in design process of final neural network. It must be noted that similar to the ANN (p,d,q) model, any aforementioned variable $e_i\,(i = t-1,\ldots,t-n)$, $\hat{L}_t$, and $z_j\,(j = t-1,\ldots,t-m)$ or set of them $\{\, e_i\,(i = t-1,\ldots,t-n)\}$ or $\{\, z_i\,(i = t-1,\ldots,t-m)\}$ may be deleted in design process of final neural network. However in the generalized hybrid ANNs/ARIMA model, in opposite of the artificial neural network (p,d,q), the linear component ($\hat{L}_t$) and original data are simultaneously applied in order to model the linear structures.

As previously mentioned, in building the autoregressive integrated moving average as well as artificial neural network models, subjective judgment of the model order as well as the model adequacy is often needed. It is possible that suboptimal models will be used in the hybrid model. For example, the current practice of Box–Jenkins methodology focuses on the low order autocorrelation. A model is considered adequate if low order autocorrelations are not significant even though significant autocorrelations of higher order still exist. This suboptimality may not affect the usefulness of the hybrid model. Granger (1989) has pointed out that for a hybrid model to produce superior forecasts, the component model should be suboptimal. In general, it has been observed that it is more effective to combine individual forecasts that are based on different information sets (Granger, 1989).

2.3.3.1. Advantages and disadvantages

Although it can be guaranteed that, the performance of the generalized hybrid ANNs/ARIMA model will not be worse than the artificial neural network (p,d,q) model and also either of the components, and a more general and more accurate model can be obtained using the above-mentioned methodology, there are not enough reasons that we can sure that the performance of the generalized hybrid ANNs/ARIMA will be also better than Zhang's hybrid model.

3. DATA SETS

Since we cannot generally demonstrate that which one of the above-mentioned methodologies is better for constructing a more appropriate and more effective hybrid model for time series forecasting, in this section, three well-known real data sets including the Wolf's sunspot data, the Canadian lynx data, and the British pound/US dollar exchange rate data are considered in order to compare the predictive capabilities of the mentioned hybrid models in practice. These time series come from different areas and have different statistical characteristics. They have been widely studied in the statistical as well as the neural network literature (Khashei & Bijari, 2010). Both linear and nonlinear models have been applied to these data sets, although more or less nonlinearities have been found in these series. Only the one-step-ahead forecasting is considered. Two performance indicators including MAE (mean absolute error) and MSE (mean squared error), which are computed from the following equations, are employed in order to measure forecasting performance of the hybrid models.

$$MAE = \frac{1}{N}\sum_{i=1}^{N}\left|e_i\right| \tag{17}$$

$$MSE = \frac{1}{N}\sum_{i=1}^{N}\left(e_i\right)^2 \tag{18}$$

3.1. The Wolf's sunspot data

The sunspot series is record of the annual activity of spots visible on the face of the sun and the number of groups into which they cluster. The sunspot data, which is considered in this investigation, contains the annual number of sunspots from 1700 to 1987, giving a total of 288 observations. The study of sunspot activity has practical importance to geophysicists, environment scientists, and climatologists. The data series is regarded as nonlinear and non-Gaussian and is often used to evaluate the effectiveness of nonlinear models (Ghiassi & Saidane, 2005). The plot of this time series (Figure 2) also suggests that there is a cyclical pattern with a mean cycle of about 11 years. The sunspot data has been extensively studied with a vast variety of linear and nonlinear time series models including ARIMA and ANNs. To assess the forecasting performance of proposed model, the sunspot data set is divided into two samples of training and testing. The training data set, 221 observations (1700- 1920), is exclusively used in order to formulate the model and then the test sample, the last 67 observations (1921- 1987), is used in order to evaluate the performance of the established model.

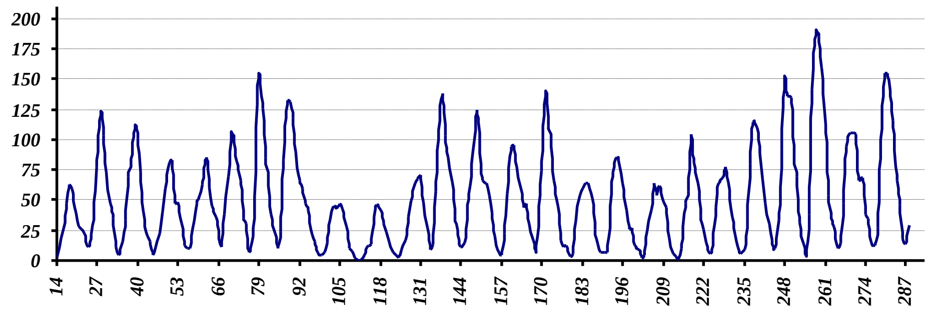

Figure 2 Annual Wolf's sunspot time series from 1700 to 1987

3.2. The Canadian lynx series

The lynx series, which is considered in this investigation, contains the number of lynx trapped per year in the Mackenzie River district of Northern Canada. The data set are plotted in Figure 3, which shows a periodicity of approximately 10 years (Stone, 2007). The data set has 114 observations, corresponding to the period of 1821–1934. It has also been extensively analyzed in the time series literature with a focus on the nonlinear modeling (Tang & Ghosal, 2007; Cornillon *et al.*, 2008) see Wong and Li (2000) for a survey. Following other studies (Zhang, 2003), the logarithms (to the base 10) of the data are used in the analysis. The training data set, 100 observations (1821- 1920), is exclusively used in order to formulate the model and then the test sample, the last 14 observations (1921- 1934), is used in order to evaluate the performance of the established model.

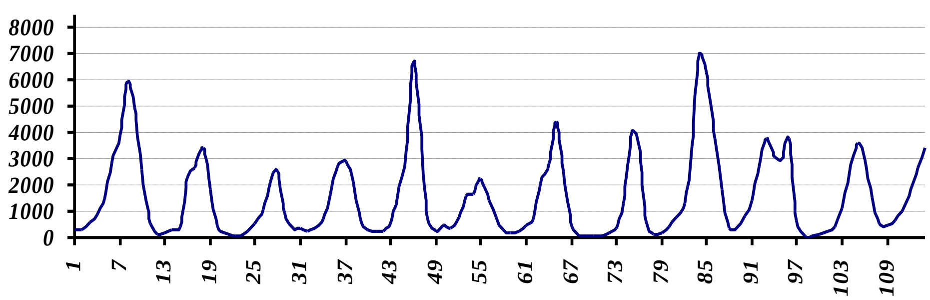

Figure 3 Annual Canadian lynx time series from 1821 to 1934

3.3. The exchange rate (British pound /US dollar)

The last data set that is considered in this investigation is the exchange rate between British pound and United States dollar. Predicting exchange rate is an important yet difficult task in international finance. Various linear and nonlinear theoretical models have been developed but few are more successful in out-of-sample forecasting than a simple random walk model. Recent applications of neural networks in this area have yielded mixed results. The data used in this paper contain the weekly observations from 1980 to 1993, giving 731 data points in the time series. The time series plot is given in Figure 4, which shows numerous changing turning points in the series. In this paper following Meese and Rogoff (1983) and Zhang (2003) and Khashei and Bijari (2010), the natural logarithmic transformed data is used in the modeling and forecasting analysis. The training data set, first 13 years (1821- 1992), is exclusively used in order to formulate the model and then the test sample, the last year (1993), is used in order to evaluate the performance of the established model.

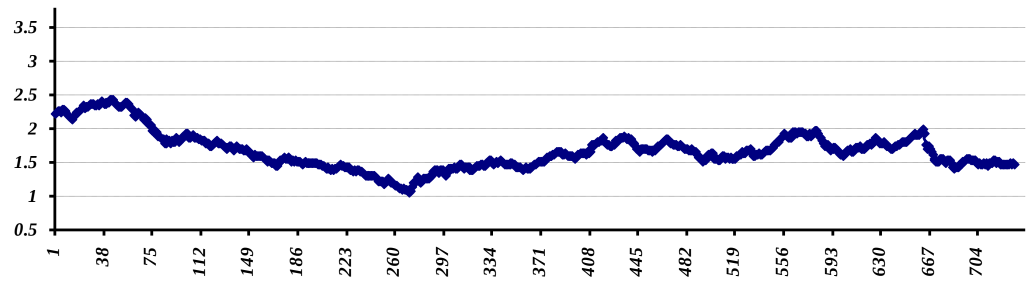

Figure 4 Weekly British pound against the United States dollar exchange rate series from 1980 to 1993

4. RESULTS

In this section, the predictive capabilities of the hybrid models including Zhang's hybrid ANNs/ARIMA, artificial neural network (p,d,q), and generalized hybrid ANNs/ARIMA are compared together and also compared with either of their components —artificial neural networks and autoregressive integrated moving average— using three above-mentioned data sets.

4.1. The Wolf's sunspot data forecasts

In the Wolf's sunspot data forecast case, according to the Akaike's information criterion (AIC), we find that a subset autoregressive model of order nine (*AR (9)*) is the most parsimonious among all ARIMA models which are also found adequate judged by the residual analysis. Many researchers such as Subba Rao and Gabr (1984), Hipel and McLeod (1994), Zhang (2003), and Khashei and Bijari (2010) have also used this model. The neural network model used is composed of four inputs, four hidden and one output neurons (in abbreviated form, $N^{(4\text{-}4\text{-}1)}$), as also employed by De Groot and Wurtz (1991), Cottrell *et al.* (1995), Zhang (2003), and Khashei and Bijari (2010). Two forecast horizons of 35 and 67 periods are used in order to assess the forecasting performance of the hybrid models and their components. The forecasting results of above-mentioned models for the sunspot data are summarized in Table 1.

Table 1 Comparison of the performance of the hybrid models and their components for sunspot data set forecasting

Model	***35 points ahead***		***67 points ahead***	
	MAE	***MSE***	***MAE***	***MSE***
Auto-Regressive Integrated Moving Average (ARIMA)	*11.319*	*216.965*	*13.033739*	*306.08217*
Artificial Neural Networks (ANNs)	*10.243*	*205.302*	*13.544365*	*351.19366*
Zhang's hybrid model	*10.831*	*186.827*	*12.780186*	*280.15956*
Artificial Neural Network (p,d,q)	*8.944*	*125.812*	*12.117994*	*234.206103*
Generalized hybrid ANNs/ARIMA	***8.847***	***129.425***	***11.446981***	***218.642153***

Results show that while applying neural networks alone can improve the forecasting accuracy over the ARIMA model in the 35-period horizon, the performance of ANNs is getting worse as time horizon extends to 67 periods. This may suggest that neither the neural network nor the ARIMA model captures all of the patterns in the data and combining two models together can be an effective way in order to overcome this limitation. However, the results of the Zhang's hybrid model show that; although, the overall forecasting errors of Zhang's hybrid model have been reduced in comparison with ARIMA and ANN, this model may also give worse predictions than either of those, in some specific situations. These results may be occurred due to the assumptions, which are considered in constructing process of the hybrid model by Zhang (2003). The obtained results of the artificial neural network (p,d,q) model (Khashei and Bijari, 2010) confirm this hypothesis that these assumptions will degenerate the performance of Zhang's hybrid model if the opposite situations occur. The artificial neural network (p,d,q) model has yielded more accurate results than Zhang's hybrid model and also both ARIMA and ANN models used separately across two different time horizons and with both error measures. However, obtained results show that more accurate results can be obtained using the generalized hybrid ANNs/ARIMA model. This model has yielded more accurate results than the artificial neural network (p,d,q), Zhang's hybrid model and also both components used in isolation across two different time horizons and with both error measures, except for MSE of the artificial neural network (p,d,q) model in the 35-period horizon.

4.2. The Canadian lynx series forecasts

In a similar fashion, we fit a subset autoregressive model of order twelve (*AR (12)*) to Canadian lynx data, according to the Akaike's information criterion (AIC). This is a parsimonious model also used by Subba Rao and Gabr (1984) and Zhang (2003), and Khashei and Bijari (2010). In addition, a neural network, which is composed of seven inputs, five hidden and one output neurons ($N^{(7\text{-}5\text{-}1)}$), has been designed to Canadian lynx data set forecast, as also employed by Zhang (2003), and Khashei and Bijari (2010). The overall forecasting results of the above-mentioned models for the last 14 years are summarized in Table 2.

Table 2 Comparison of the performance of the hybrid models and their components for Canadian lynx data set forecasting

Model	*MAE*	*MSE*
Auto-Regressive Integrated Moving Average (ARIMA)	*0.112255*	*0.020486*
Artificial Neural Networks (ANNs)	*0.112109*	*0.020466*
Zhang's hybrid model	*0.103972*	*0.017233*
Artificial Neural Network (p,d,q)	*0.089625*	*0.013609*
Generalized hybrid ANNs/ARIMA	***0.085055***	***0.00999***

Numerical results show that the used neural network gives slightly better forecasts than the ARIMA model and the Zhang's hybrid model, significantly outperform the both of them. However, according to the previous case, the obtained results of the artificial neural network (p,d,q) model are better than Zhang's hybrid model and the obtained results of the generalized hybrid ANNs/ARIMA model are better than the artificial neural network (p,d,q) model in both error measures.

4.3. The exchange rate (British pound /US dollar) forecasts

With the exchange rate data set and according to the Akaike's information criterion (AIC), the best linear ARIMA model is found to be the simple random walk model: $y_t = y_{t-1} + \varepsilon_t$. This is the same finding suggested by many studies in the exchange rate literature that a simple random walk is the dominant linear model. They claim that the evolution of any exchange rate follows the theory of efficient market hypothesis (EMH) (Timmermann & Granger, 2004). According to this hypothesis, the best prediction value for tomorrow's exchange rate is the current value of the exchange rate and the actual exchange rate follows a random walk. A neural network, which is composed of seven inputs, six hidden and one output neurons ($N^{(7\text{-}6\text{-}1)}$) is designed in order to model the nonlinear patterns, as also employed by others (Zhang, 2003; Khashei & Bijari, 2010). Three time horizons of 1, 6 and 12 months are used in order to assess the forecasting performance of models. The forecasting results of above-mentioned models for the exchange rate data are summarized in Table 3.

Table 3 Comparison of the performance of the proposed model with those of other forecasting models (exchange rate data)*

Model	***1 month***		***6 month***		***12 month***	
	MAE	***MSE***	***MAE***	***MSE***	***MAE***	***MSE***
Auto-Regressive Integrated Moving Average	*0.005016*	*3.68493*	*0.0060447*	*5.65747*	*0.0053579*	*4.52977*
Artificial Neural Networks (ANNs)	*0.004218*	*2.76375*	*0.0059458*	*5.71096*	*0.0052513*	*4.52657*
Zhang's hybrid model	*0.004146*	*2.67259*	*0.0058823*	*5.65507*	*0.0051212*	*4.35907*
Artificial Neural Network (p,d,q)	*0.004001*	*2.60937*	*0.0054440*	*4.31643*	*0.0051069*	*3.76399*
Generalized hybrid ANNs/ARIMA	***0.003972***	***2.39915***	***0.0053361***	***4.27822***	***0.0049691***	***3.64774***

* Note: All MSE values should be multiplied by 10^{-5}.

In the exchange rate data set forecasting, similar to the previous section, the performance of the generalized hybrid ANNs/ARIMA model is better than the artificial neural network (p,d,q) model, the performance of the artificial neural network (p,d,q) model is better than Zhang's hybrid model, and the performance of the Zhang's hybrid model is better than either of the components across three different time horizons and with both error measures.

5. CONCLUSIONS

In this paper, the predictive capabilities of three different hybrid linear and nonlinear models in which the autoregressive integrated moving average (ARIMA) as linear model is combined to the multilayer perceptron (MLP) as nonlinear model are compared together for time series forecasting. These models include Zhang's hybrid ANNs/ARIMA, the artificial neural network (p,d,q), and the generalized hybrid ANNs/ARIMA models. Some general results obtained from comparing these models together are as follows:

1- It can be generally guaranteed that the performance of the Zhang's hybrid model will not be worse than autoregressive integrated moving average (ARIMA) model.

2- It cannot be generally guaranteed that the performance of the Zhang's hybrid model will not be worse than the multilayer perceptron (MLP) model.

3- It can be generally guaranteed that the performance of the artificial neural network (p,d,q) model will not be worse than either of the components including autoregressive integrated moving average (ARIMA) and multilayer perceptron (MLP) models.

4- It cannot be generally guaranteed that the performance of the artificial neural network (p,d,q) model will not be worse than the Zhang's hybrid model.

5- It can be generally guaranteed that the performance of the generalized hybrid ANNs/ARIMA model will not be worse than either of the components including autoregressive integrated moving average (ARIMA) and multilayer perceptron (MLP) models.

6- It can be generally guaranteed that the performance of the generalized hybrid ANNs/ARIMA model will not be worse than the artificial neural network (p,d,q) model.

7- It cannot be generally guaranteed that the performance of the generalized hybrid ANNs/ARIMA model will not be worse than the Zhang's hybrid model.

Since, it cannot be generally demonstrated that the obtained results of which one of these models is more accurate, the predictive capabilities of the above-mentioned hybrid models are practically compared together. Empirical results with three well-known real data sets including the Wolf's sunspot data, the Canadian lynx data, and the British pound against the United States dollar exchange rate data, indicate that while all of these methodologies can be an effective way to improve forecasting accuracy achieved by either of components used separately, the generalized hybrid ANNs/ARIMA model is more accurate and perform better than artificial neural network (p,d,q) and Zhang's hybrid ANNs/ARIMA models.

ACKNOWLEDGEMENTS

The authors wish to express their gratitude to, Seyed Reza Hejazi, assistant professor of industrial engineering, Isfahan University of Technology, who greatly helped us.

REFERENCES

[1] Armano G., Marchesi M., Murru A. (2005), A hybrid genetic-neural architecture for stock indexes forecasting; *Information Sciences* 170; 3–33.

[2] Aryal D.R., Yao-Wu W. (2003), Neural network Forecasting of the production level of Chinese construction industry; *Journal of Comparative International Management* 6(2); 45–64.

[3] Balkin S.D., Ord J.K. (2000), Automatic neural network modeling for univariate time series; *International Journal of Forecasting* 16; 509–515.

[4] Bates J.M., Granger W.J. (1969), The combination of forecasts; *Operation Research* 20; 451–468.

[5] Berardi V.L., Zhang G.P. (2003), An empirical investigation of bias and variance in time series forecasting: modeling considerations and error evaluation; *IEEE Transactions on Neural Networks* 14(3); 668–679.

[6] Bollerslev T. (1986), Generalized autoregressive conditional heteroscedasticity; *Journal of Econometrics* 31; 307–327.

[7] Box P., Jenkins G.M. (1976), Time Series Analysis: Forecasting and Control; Holden-day Inc, San Francisco, CA.

[8] Brace M.C., Schmidt J., Hadlin M. (1991), Comparison of the forecasting accuracy of neural networks with other established techniques; *Proceedings of the First Forum on Application for weight elimination, IEEE Transactions on Neural Networks of Neural Networks to Power Systems*; Seattle, WA, 31– 35.

[9] Chakraborty K., Mehrotra K., Mohan C.K., Ranka S. (1992), Forecasting the behavior of multivariate time series using neural networks; *Neural Networks* 5; 961–970.

[10] Chappel D., Padmore J., Mistry P., Ellis C. (1996), A threshold model for the French franc/Deutschmark exchange rate; *Journal of Forecasting* 15(3); 155–164.

[11] Chen K.Y., Wang C.H. (2007), A hybrid SARIMA and support vector machines in forecasting the production values of the machinery industry in Taiwan; *Expert Systems with Applications* 32; 254–264.

[12] Chen Y., Yang B., Dong J., Abraham A. (2005), Time-series forecasting using flexible neural tree model; *Information Sciences* 174(3–4); 219–235.

[13] Clemen R. (1989), Combining forecasts: a review and annotated bibliography with discussion; *International Journal of Forecasting* 5; 559–608.

[14] Cornillon P., Imam W., Matzner E. (2008), Forecasting time series using principal component analysis with respect to instrumental variables; *Computational Statistics & Data Analysis* 52; 1269–1280.

[15] Cottrell M., Girard B., Girard Y., Mangeas M., Muller C. (1995), Neural modeling for time series: a statistical stepwise method for weight elimination; *IEEE Transactions on Neural Networks* 6(6); 1355–1364.

[16] De Groot C., Wurtz D. (1991), Analysis of univariate time series with connectionist nets: a case study of two classical examples; *Neurocomputing* 3; 177–192.

[17] Denton J.W. (1995), How good are neural networks for causal forecasting?; *The Journal of Business Forecasting* 14(2); 17–20.

[18] Engle R.F. (1982), Autoregressive conditional heteroskedasticity with estimates of the variance of UK inflation; *Econometrica* 50; 987–1008.

[19] Fishwick P.A. (1989), Neural network models in simulation: A comparison with traditional modeling approaches; *Proceedings of Winter Simulation Conference*, Washington D.C.; 702–710.

[20] Foster W.R., Collopy F., Ungar L.H. (1992), Neural network forecasting of short, noisy time series; *Computers and Chemical Engineering* 16(4); 293– 297.

[21] Ghiassi M., Saidane H. (2005), A dynamic architecture for artificial neural networks; *Neurocomputing* 63; 397–413.

[22] Ginzburg I., Horn D. (1994), Combined neural networks for time series analysis; *Adv. Neural Inf. Process. Systems* 6; 224–231.

[23] Giordano F., Rocca M., Perna C. (2007), Forecasting nonlinear time series with neural network sieve bootstrap; *Computational Statistics and Data Analysis* 51; 3871–3884.

[24] Goh W.Y., Lim C.P., Peh K.K. (2003), Predicting drug dissolution profiles with an ensemble of boosted neural networks: a time series approach; *IEEE Transactions on Neural Networks* 14(2); 459–463.

[25] Granger C.W.J. (1989), Combining forecasts—Twenty years later; *Journal of Forecasting* 8; 167–173.

[26] Granger C.W.J., Anderson A.P. (1978), An Introduction to Bilinear Time Series Models; Vandenhoeck and Ruprecht, Go¨ttingen.

[27] Hann T.H., Steurer E. (1996), Much ado about nothing? Exchange rate forecasting: neural networks vs. linear models using monthly and weekly data; *Neurocomputing* 10; 323–339.

[28] Haseyama M., Kitajima H. (2001), An ARMA order selection method with fuzzy reasoning; *Signal Process* 81; 1331–1335.

[29] Hipel K.W., McLeod A.I. (1994), Time Series Modelling of Water Resources and Environmental Systems; Amsterdam, Elsevier.

[30] Hsieh D.A. (1991), Chaos and nonlinear dynamics: application to financial markets; *Journal of Finance* 46; 1839–1877.

[31] Hurvich C.M., Tsai C.-L. (1989), Regression and time series model selection in small samples; *Biometrica* 76 (2); 297–307.

[32] Hwang H.B. (2001), Insights into neural-network forecasting time series corresponding to ARMA (*p; q*) structures; *Omega* 29; 273–289.

[33] Jain A., Kumar A.M. (2007), Hybrid neural network models for hydrologic time series forecasting; *Applied Soft Computing* 7; 585– 592.

[34] Jones R.H. (1975), Fitting autoregressions; *J. Amer. Statist. Assoc.* 70 (351); 590–592.

[35] Khashei M., Bijari M. (2010), An artificial neural network (*p*, *d*, *q*) model for time series forecasting; *Expert Systems with Applications* 37; 479–489.

[36] Khashei M., Bijari M. (2011), A novel hybridization of artificial neural networks and ARIMA models for time series forecasting; *Applied Soft Computing* 11; 2664–2675.

[37] Khashei M., Bijari M., Raissi GH.A. (2009), Improvement of Auto-Regressive Integrated Moving Average Models Using Fuzzy Logic and Artificial Neural Networks (ANNs); *Neurocomputing* 72; 956– 967.

[38] Khashei M., Hejazi S.R., Bijari M. (2008), A new hybrid artificial neural networks and fuzzy regression model for time series forecasting; *Fuzzy Sets and Systems* 159; 769–786.

[39] Kim H., Shin K. (2007), A hybrid approach based on neural networks and genetic algorithms for detecting temporal patterns in stock markets; *Applied Soft Computing* 7; 569–576.

[40] Lapedes A., Farber R. (1987), Nonlinear signal processing using neural networks: prediction and system modeling; Technical Report LAUR-87-2662; Los Alamos National Laboratory, Los Alamos, NM.

[41] Ljung L. (1987), System Identification Theory for the User; Prentice-Hall, Englewood Cliffs, NJ.

[42] Luxhoj J.T., Riis J.O., Stensballe B. (1996), A hybrid econometric-neural network modeling approach for sales forecasting; *Int. J. Prod. Econ.* 43; 175–192.

[43] Makridakis S. (1989), Why combining works?; *International Journal of Forecasting* 5; 601–603.

[44] Meese R.A., Rogoff K. (1983), Empirical exchange rate models of the seventies: do they fit out of samples?; *J. Int. Econ.* 14; 3–24.

[45] Minerva T., Poli I. (2001), Building ARMA models with genetic algorithms; *Lecture Notes in Computer Science* 2037; 335–342.

[46] Mizrach B. (1992), Multivariate nearest-neighbor forecasts of EMS exchange rates'; *Journal of Applied Econometrics* 7; 151–164.

[47] Ong C.-S., Huang J.-J., Tzeng G.-H. (2005), Model identification of ARIMA family using genetic algorithms; *Appl. Math. Comput.* 164(3); 885–912.

[48] Pai P.F., Lin C.S. (2005), A hybrid ARIMA and support vector machines model in stock price forecasting; *Omega* 33; 497–505.

[49] Panda C., Narasimhan V. (2007), Forecasting exchange rate better with artificial neural network; *Journal of Policy Modeling* 29; 227–236.

[50] Pelikan E., de Groot C., Wurtz D. (1992), Power consumption in West-Bohemia: improved forecasts with decorrelating connectionist networks; *Neural Network World* 2; 701–712.

[51] Poli I., Jones R.D. (1994), A neural net model for prediction; *Journal of American Statistical Association* 89; 117–121.

[52] Ragulskis M., Lukoseviciute K. (2009), Non-uniform attractor embedding for time series forecasting by fuzzy inference systems; *Neurocomputing* 72, 2618–2626.

[53] Reid M.J. (1968), Combining three estimates of gross domestic product; *Economica* 35; 431–444.

[54] Santos A., da Costa Jr N., Coelho L. (2007), Computational intelligence approaches and linear models in case studies of forecasting exchange rates; *Expert Systems with Applications* 33; 816–823.

[55] Shibata R. (1976), Selection of the order of an autoregressive model by Akaike's information criterion*;* *Biometrika AC*-63 (1); 117–126.

[56] Stone L., He D. (2007), Chaotic oscillations and cycles in multi-trophic ecological systems; *Journal of Theoretical Biology* 248; 382–390.

[57] Subba Rao T., Sabr M.M. (1984), An Introduction to Bispectral Analysis and Bilinear Time Series Models; *Lecture Notes in Statistics* 24; Springer-Verlag, New York.

[58] Tang Y., Ghosal S. (2007), A consistent nonparametric Bayesian procedure for estimating autoregressive conditional densities; *Computational Statistics & Data Analysis* 51; 4424–4437.

[59] Tang Z., Almeida C., Fishwick P.A. (1991), Time series forecasting using neural networks vs. Box-Jenkins methodology; *Simulation* 57(5); 303–310.

[60] Tang Z., Fishwick P.A. (1993), Feedforward neural nets as models for time series forecasting; *ORSA Journal on Computing* 5(4); 374–385.

[61] Taskaya T., Ahmad K. (2005), Are ARIMA neural network hybrids better than single models?; *Proceedings of International Joint Conference on Neural Networks (IJCNN 2005)*; July 31–August 4, Canada.

[62] Taskaya T., Casey M. C. (2005), A comparative study of autoregressive neural network hybrids; *Neural Networks* 18; 781–789.

[63] Timmermann A., Granger C.W.J. (2004), Efficient market hypothesis and forecasting; *Int. J. Forecasting* 20; 15–27.

[64] Tong H., Lim K.S. (1980), Threshold autoregressive, limit cycles and cyclical data; *Journal of the Royal Statistical Society Series B* 42(3); 245–292.

[65] Tsaih R., Hsu Y., Lai C.C. (1998), Forecasting S&P 500 stock index futures with a hybrid AI system; *Decision Support Systems* 23; 161–174.

[66] Tseng F.M., Yu H.C., Tzeng G.H. (2002), Combining neural network model with seasonal time series ARIMA model; *Technological Forecasting & Social Change* 69; 71–87.

[67] Voort M.V.D., Dougherty M., Watson S. (1996), Combining Kohonen maps with ARIMA time series models to forecast traffic flow; *Transportation Research Part C: Emerging Technologies* 4; 307–318.

[68] Wedding D.K., Cios K.J. (1996), Time series forecasting by combining RBF networks, certainty factors, and the Box–Jenkins model; *Neurocomputing* 10; 149–168.

[69] Weigend A.S., Gershenfeld N.A. (1993), Time Series Prediction: Forecasting the Future and Understanding the Past; Addison-Wesley, Reading, MA.

[70] Wold H. (1938), A Study in the Analysis of Stationary Time Series; Almgrist & Wiksell, Stockholm.

[71] Wong C.S., Li W.K. (2000), On a mixture autoregressive model; *J. Roy. Statist. Soc. Ser. B* 62(1); 91–115.

[72] Yu L., Wang S., Lai K.K. (2005), A novel nonlinear ensemble forecasting model incorporating GLAR and ANN for foreign exchange rates; *Computers and Operations Research* 32; 2523–2541.

[73] Yule G.U. (1926), Why do we sometimes get nonsense-correlations between time series? A study in sampling and the nature of time series; *J. R. Statist. Soc.* 89; 1–64.

[74] Zhang G., Patuwo B.E., Hu M.Y. (1998), Forecasting with artificial neural networks: The state of the art; *International Journal of Forecasting* 14; 35– 62.

[75] Zhang G.P. (2007), A neural network ensemble method with jittered training data for time series forecasting; *Information Sciences* 177; 5329–5346.

[76] Zhang G.P. (2003), Time series forecasting using a hybrid ARIMA and neural network model; *Neurocomputing* 50; 159–175.

[77] Zhou Z.J., Hu C.H. (2008), An effective hybrid approach based on grey and ARMA for forecasting gyro drift, Chaos; *Solitons and Fractals* 35; 525–529.

[78] Zou H.F., Xia G.P., Yang F.T., Wang H.Y. (2007), An investigation and comparison of artificial neural network and time series models for Chinese food grain price forecasting; *Neurocomputing* 70; 2913–2923.