ELSEVIER

NEUROCOMPUTING

Neurocomputing 50 (2003) 159–175

www.elsevier.com/locate/neucom

Time series forecasting using a hybrid ARIMA and neural network model

G. Peter Zhang*

Department of Management, J. Mack Robinson College of Business, Georgia State University, University Plaza, Atlanta, GA 30303, USA

Received 16 July 1999; accepted 23 November 2001

Abstract

Autoregressive integrated moving average (ARIMA) is one of the popular linear models in time series forecasting during the past three decades. Recent research activities in forecasting with artificial neural networks (ANNs) suggest that ANNs can be a promising alternative to the traditional linear methods. ARIMA models and ANNs are often compared with mixed conclusions in terms of the superiority in forecasting performance. In this paper, a hybrid methodology that combines both ARIMA and ANN models is proposed to take advantage of the unique strength of ARIMA and ANN models in linear and nonlinear modeling. Experimental results with real data sets indicate that the combined model can be an effective way to improve forecasting accuracy achieved by either of the models used separately.

Keywords: ARIMA; Box–Jenkins methodology; Artificial neural networks; Time series forecasting; Combined forecast

1. Introduction

Time series forecasting is an important area of forecasting in which past observations of the same variable are collected and analyzed to develop a model describing the underlying relationship. The model is then used to extrapolate the time series into the future. 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. Much effort has been devoted over the past several decades to the development and improvement of time series forecasting models.

* Tel.: +1-404-651-4065; fax: +1-404-651-3498.
E-mail address: gpzhang@gsu.edu (G.P. Zhang).

PII: S0925-2312(01)00702-0

One of the most important and widely used time series models is the autoregressive integrated moving average (ARIMA) model. The popularity of the ARIMA model is due to its statistical properties as well as the well-known Box–Jenkins methodology [2] in the model building process. In addition, various exponential smoothing models can be implemented by ARIMA models [26]. Although ARIMA models are quite flexible in that they can represent several different types of time series, i.e., pure autoregressive (AR), pure moving average (MA) and combined AR and MA (ARMA) series, their major limitation is the pre-assumed linear form of the model. That is, a linear correlation structure is assumed among the time series values and therefore, no nonlinear patterns can be captured by the ARIMA model. The approximation of linear models to complex real-world problem is not always satisfactory.

Recently, artificial neural networks (ANNs) have been extensively studied and used in time series forecasting. Zhang et al. [43] presented a recent review in this area. The major advantage of neural networks is their flexible nonlinear modeling capability. With ANNs, there is no need to specify a particular model form. Rather, the model is adaptively formed based on the features presented from the data. This data-driven approach is suitable for many empirical data sets where no theoretical guidance is available to suggest an appropriate data generating process.

In this paper, we propose a hybrid approach to time series forecasting using both ARIMA and ANN models. The motivation of the hybrid model comes 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 or whether one particular method is more effective than the other in out-of-sample forecasting. Thus, it is difficult for forecasters to choose the right technique for their unique situations. Typically, a number of different models are tried and the one with the most accurate result is selected. However, the final selected model is not necessarily the best for future uses due to many potential influencing factors such as sampling variation, model uncertainty, and structure change. By combining different methods, the problem of model selection can be eased with little extra effort. Second, real-world time series are rarely pure linear or nonlinear. They often contain both linear and nonlinear patterns. If this is the case, then neither ARIMA nor NNs can be adequate in modeling and forecasting time series since the ARIMA model cannot deal with nonlinear relationships while the neural network model alone is not able to handle both linear and nonlinear patterns equally well. Hence, by combining ARIMA with ANN models, complex autocorrelation structures in the data can be modeled more accurately. Third, it is almost universally agreed in the forecasting literature that no single method is best in every situation [4,19,23]. This is largely 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. For example, in the literature of time series forecasting with neural networks, most studies [34,36,37,44–47] use the ARIMA models as the benchmark to test the effectiveness of the ANN model with mixed results. Many empirical studies including several large-scale forecasting competitions suggest that by combining several different models, forecasting accuracy can often be improved over the individual model without the need to find the "true" or "best" model [6,23,24,28]. Therefore, combining different models can increase the chance to capture different patterns in the data and improve forecasting performance. Several empirical studies have already suggested that by combining several different models, forecasting accuracy can often be improved over the individual model. In addition, the combined model is more robust with regard to the possible structure change in the data.

Using hybrid model or combining several models has become a common practice to improve the forecasting accuracy since the well-known M-competition [23] in which combination of forecasts from more than one model often leads to improved forecasting performance. The literature on this topic has expanded dramatically since the early work of Reid [32] and Bates and Granger [1]. Clemen [6] 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 forecasts [22,28,29,40]. In neural network forecasting research, a number of combining schemes have been proposed. Wedding and Cios [39] described a combining methodology using radial basis function networks and the Box–Jenkins models. Luxhoj et al. [21] presented a hybrid econometric and ANN approach for sales forecasting. Pelikan et al. [30] and Ginzburg and Horn [13] proposed to combine several feedforward neural networks to improve time series forecasting accuracy.

The rest of the paper is organized as follows. In the next section, we review the ARIMA and ANN modeling approaches to time series forecasting. The hybrid methodology is introduced in Section 3. Empirical results from three real data sets are reported in Section 4. Section 5 contains the concluding remarks.

2. Time series forecasting models

There are several different approaches to time series modeling. Traditional statistical models including moving average, exponential smoothing, and ARIMA are linear in that predictions of the future values are constrained to be linear functions of past observations. Because of their relative simplicity in understanding and implementation, linear models have been the main research focuses and applied tools during the past few decades. To overcome the restriction of the linear models and to account for certain nonlinear patterns observed in real problems, several classes of nonlinear models have been proposed in the literature. These include the bilinear model [14], the threshold autoregressive (TAR) model [38], and the autoregressive conditional heteroscedastic (ARCH) model [11]. Although some improvement has been noticed with these nonlinear models, the gain of using them to general forecasting problems is limited [8]. Because these models are developed for specific nonlinear patterns, they are not capable of modeling other types of nonlinearity in time series. More recently, artificial neural networks have been suggested as an alternative to time series forecasting. The main strength of the ANNs is their flexible nonlinear modeling capability. In this section, we focus on the basic principles and modeling process of the ARIMA and ANN models.

2.1. The ARIMA model

In an autoregressive integrated moving average 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 generate the time series has the form

$$y_t = \theta_0 + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{t-q}, \quad (1)$$

where y_t and ε_t are the actual value and random error at time period t, respectively; ϕ_i $(i = 1, 2, \ldots, p)$ and θ_j $(j = 0, 1, 2, \ldots, q)$ are model parameters. p and q are integers and often referred to as orders of the model. Random errors, ε_t, are assumed to be independently and identically distributed with a mean of zero and a constant variance of σ^2.

Eq. (1) entails several important special cases of the ARIMA family of models. If $q = 0$, then (1) becomes an AR model of order p. When $p = 0$, the model reduces to an MA model of order q. One central task of the ARIMA model building is to determine the appropriate model order (p, q).

Based on the earlier work of Yule [42] and Wold [41], Box and Jenkins [2] 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 ARIMA 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 [2] proposed to use the autocorrelation function and the partial autocorrelation function of the sample data as the basic tools to identify the order of the ARIMA model.

In the identification step, data transformation is often needed to make the time series stationary. Stationarity is a necessary condition in building an ARIMA model that is useful for forecasting. A stationary time series has the property that its statistical characteristics such as the mean and the autocorrelation structure are constant over time. When the observed time series presents trend and heteroscedasticity, differencing and power transformation are often applied to the data to remove the trend and stabilize the variance before an ARIMA model can be fitted.

Once a tentative model is specified, estimation of the model parameters is straightforward. The parameters are estimated such that an overall measure of errors is minimized. This can be done with a nonlinear optimization procedure.

The last step of model building is the diagnostic checking of model adequacy. This is basically to check if the model assumptions about the errors, ε_t, are satisfied. 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 is again 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.

2.2. The ANN approach to time series modeling

When the linear restriction of the model form is relaxed, the possible number of nonlinear structures that can be used to describe and forecasting a time series is enormous. A good nonlinear model should be "general enough to capture some of the nonlinear phenomena in the data" [8]. Artificial neural networks are one of such models that are able to approximate various nonlinearities in the data.

ANNs are flexible computing frameworks for modeling a broad range of nonlinear problems. One significant advantage of the ANN models over other classes of nonlinear model is that ANNs are universal approximators which can approximate a large class of functions with a high degree of accuracy. 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 feedforward network is the most widely used model form for time series modeling and forecasting [43]. The model is characterized by a network of three layers of simple processing units connected by acyclic links. The relationship between the output (y_t) and the inputs ($y_{t-1}, y_{t-2}, \ldots, y_{t-p}$) has the following mathematical representation:

$$y_t = \alpha_0 + \sum_{j=1}^{q} \alpha_j g\left(\beta_{0j} + \sum_{i=1}^{p} \beta_{ij} y_{t-i}\right) + \varepsilon_t, \tag{2}$$

where α_j ($j = 0, 1, 2, \ldots, q$) and β_{ij} ($i = 0, 1, 2, \ldots, p$; $j = 1, 2, \ldots, q$) are the model parameters often called the connection weights; p is the number of input nodes and q is the number of hidden nodes. The logistic function is often used as the hidden layer transfer function, that is,

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

Hence, the ANN model of (2) in fact performs a nonlinear functional mapping from the past observations ($y_{t-1}, y_{t-2}, \ldots, y_{t-p}$) to the future value y_t, i.e.,

$$y_t = f(y_{t-1}, y_{t-2}, \ldots, y_{t-p}, w) + \varepsilon_t, \tag{4}$$

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. Note that expression (2) implies one output node in the output layer which is typically used for one-step-ahead forecasting.

The simple network given by (2) is surprisingly powerful in that it is able to approximate arbitrary function as the number of hidden nodes q is sufficiently large [17]. 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 neural network modeling process. An overfitted model has a good fit to the sample used for model building but has poor generalization ability for data out of the sample. 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, 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. However, there is no theory that can be used to guide the selection of p. Hence, experiments are often conducted to select an appropriate p as well as q.

Once a network structure (p, q) is specified, the network is ready for training—a process of parameter estimation. As in ARIMA model building, the parameters are estimated such that an overall accuracy criterion such as the mean squared error is minimized. This is again done with some efficient nonlinear optimization algorithms other than the basic backpropagation training algorithm. One of these is generalized reduced gradient (GRG2), a general purpose nonlinear optimizer [18]. A GRG2-based training system is used in this study [35].

The estimated model is usually evaluated using a separate hold-out sample that is not exposed to the training process. This practice is different from that in ARIMA model building where one sample is typically used for model identification, estimation and evaluation. The reason lies in the fact that the general (linear) form of the ARIMA model is pre-specified and then the order of the model is estimated from the data. The standard statistical paradigm assumes that under stationary condition, the model best fitted to the historical data is also the optimum model for forecasting [12]. With ANNs, the (nonlinear) model form as well as the order of the model must be estimated from the data. It is, therefore, more likely for an ANN model to overfit the data.

There are some similarities between ARIMA and ANN models. Both of them include a rich class of different models with different model orders. Data transformation is often necessary to get best results. A relatively large sample is required in order to build a successful model. The iterative experimental nature is common to their modeling processes and the subjective judgement is sometimes needed in implementing the model. Because of the potential overfitting effect with both models, parsimony is often a guiding principle in choosing an appropriate model for forecasting.

3. The hybrid methodology

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. For example, using simulated data, Denton [10] showed that when there are outliers or multicollinearity in the data, neural networks can significantly outperform linear regression models. Markham and Rakes [25] also found that the performance of ANNs for linear regression problems depends on the sample size and noise level. 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.

It may be reasonable to consider a time series to be composed of a linear autocorrelation structure and a nonlinear component. That is,

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

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, then the residuals from the linear model will contain only the nonlinear relationship. 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 (2). Residuals are important in diagnosis of the sufficiency of linear models. A linear model is not sufficient if there are still linear correlation structures left in the residuals. However, residual analysis is not able to detect any nonlinear patterns in the data. In fact, there is currently no general diagnostic statistics for nonlinear autocorrelation relationships. Therefore, even if a model has passed diagnostic checking, the model may still not be adequate in that nonlinear relationships have not been appropriately modeled. Any significant nonlinear pattern in the residuals will indicate the limitation of the ARIMA. By modeling residuals using ANNs, nonlinear relationships can be discovered. With n input nodes, the ANN model for the residuals will be

$$e_t = f(e_{t-1}, e_{t-2}, \ldots, e_{t-n}) + \varepsilon_t, \tag{7}$$

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

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

In summary, the proposed methodology of the hybrid system consists of two steps. In the first step, an ARIMA model is used to analyze the linear part of the problem. In the second step, a neural network model is developed to model the residuals from the ARIMA model. Since the ARIMA model cannot capture the nonlinear structure of the data, the residuals of linear model will contain information about the nonlinearity. The results from the neural network can be used as predictions of the error terms for the ARIMA model. 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.

As previously mentioned, in building ARIMA as well as ANN models, subjective judgement of the model order as well as the model adequacy is often needed. It is possible that suboptimal models may be used in the hybrid method. 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 [15] 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 [15,31].

4. Empirical results

4.1. Data sets

Three well-known data sets—the Wolf's sunspot data, the Canadian lynx data, and the British pound/US dollar exchange rate data—are used in this study to demonstrate the effectiveness of the hybrid method. 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. Both linear and nonlinear models have been applied to these data sets, although more or less nonlinearities have been found in these series.

The sunspot data we consider 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 [16]. The data series is regarded as nonlinear and non-Gaussian and is often used to evaluate the effectiveness of nonlinear models. The plot of this time series (see Fig. 1) also suggests that there is a cyclical pattern with a mean cycle of about 11 years. The sunspot data

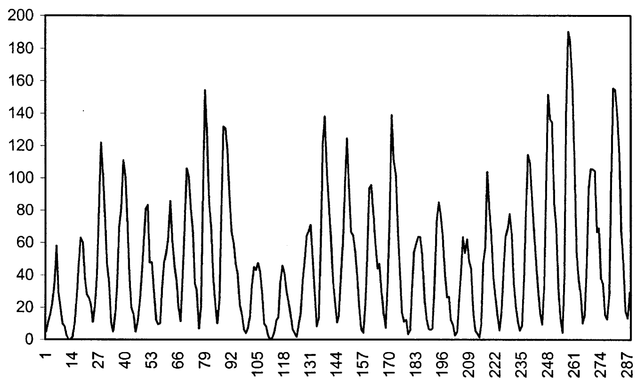

Fig. 1. Sunspot series (1700–1987).

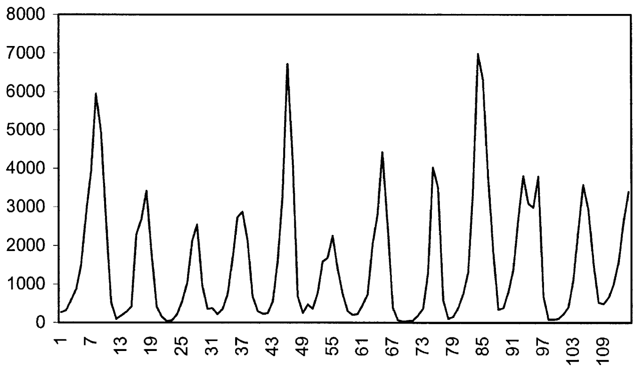

Fig. 2. Canadian lynx data series (1821–1934).

Fig. 3. Weekly BP/USD exchange rate series (1980–1993).

have been extensively studied with a vast variety of linear and nonlinear time series models including ARIMA and ANNs.

The lynx series contains the number of lynx trapped per year in the Mackenzie River district of Northern Canada. The data are plotted in Fig. 2, which shows a periodicity of approximately 10 years. 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. Following other studies [3,33], the logarithms (to the base 10) of the data are used in the analysis.

The last data set is the exchange rate between British pound and US 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 Fig. 3, which shows numerous changing turning

Table 1
Sample compositions in three data sets

Series	Sample size	Training set (size)	Test set (size)
Sunspot	288	1700–1920 (221)	1921–1987 (67)
Lynx	114	1821–1920 (100)	1921–1934 (14)
Exchange rate	731	1980–1992 (679)	1993 (52)

Table 2
Forecasting comparison for sunspot data

	35 points ahead		67 points ahead	
	MSE	MAD	MSE	MAD
ARIMA	216.965	11.319	306.08217	13.033739
ANN	205.302	10.243	351.19366	13.544365
Hybrid	186.827	10.831	280.15956	12.780186

points in the series. Following Meese and Rogoff [27], we use the natural logarithmic transformed data in the modeling and forecasting analysis.

To assess the forecasting performance of different models, each data set is divided into two samples of training and testing. The training data set is used exclusively for model development and then the test sample is used to evaluate the established model. The data compositions for the three data sets are given in Table 1.

4.2. Results

In this study, all ARIMA modeling is implemented via SAS/ETS system while neural network models are built using the GRG2-based training system mentioned earlier. Only the one-step-ahead forecasting is considered. The mean squared error (MSE) and mean absolute deviation (MAD) are selected to be the forecasting accuracy measures.

Table 2 gives the forecasting results for the sunspot data. A subset autoregressive model of order 9 has been found to be the most parsimonious among all ARIMA models that are also found adequate judged by the residual analysis. This model has also been used by many researchers such as Subba Rao and Gabr [33] and Hipel and McLeod [16]. The neural model used is a $4 \times 4 \times 1$ network as also employed by De Groot and Wurtz [9] and Cottrell et al. [7]. Two forecast horizons of 35 and 67 periods are used. 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. The results of the hybrid model show that by combining two models together, the overall forecasting errors can be significantly reduced except for the 35-period forecasting with the MAD measure where ANN is slightly better. In terms of MSE, the percentage improvements of the hybrid model over the ARIMA and ANN for 35-period forecasts are 16.13% and 9.89%, respectively. The comparison between the

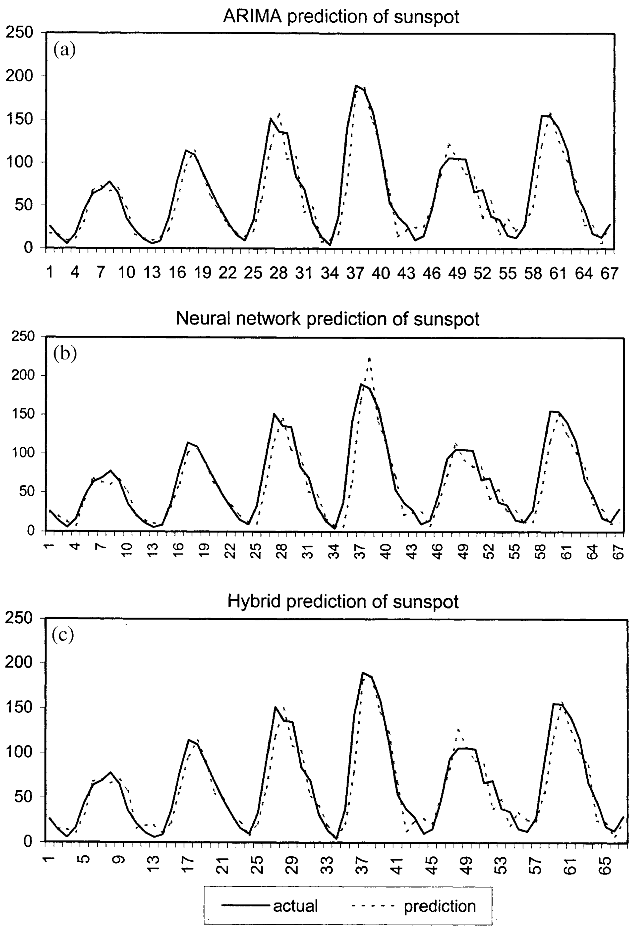

Fig. 4. (a) ARIMA prediction of sunspot. (b) Neural network prediction of sunspot and (c) Hybrid prediction of sunspot.

actual value and the forecast value for the 67 points out-of-sample is given in Fig. 4. Although at some data points, the hybrid model gives worse predictions than either ARIMA or ANN forecasts, its overall forecasting capability is improved.

In a similar fashion, we have fitted to Canadian lynx data with a subset AR model of order 12. This is a parsimonious model also used by Subba Rao and Gabr [33] and others. The overall forecasting results for the last 14 years are summarized in Table 3. A neural network structure of $7 \times 5 \times 1$ gives slightly better forecasts than the ARIMA model. Applying the hybrid method, we find an 18.87% (18.76%) decrease in MSE over ARIMA (ANN). With MAD, the improvement of the hybrid model over the ARIMA and ANN are 7.97% and 7.83%, respectively. Fig. 5 gives the actual vs. forecast values with individual models of ANN and ARIMA as well as the combined model.

Table 3
Lynx forecasting results

	MSE	MAD
ARIMA	0.020486	0.112255
ANN	0.020466	0.112109
Hybrid	0.017233	0.103972

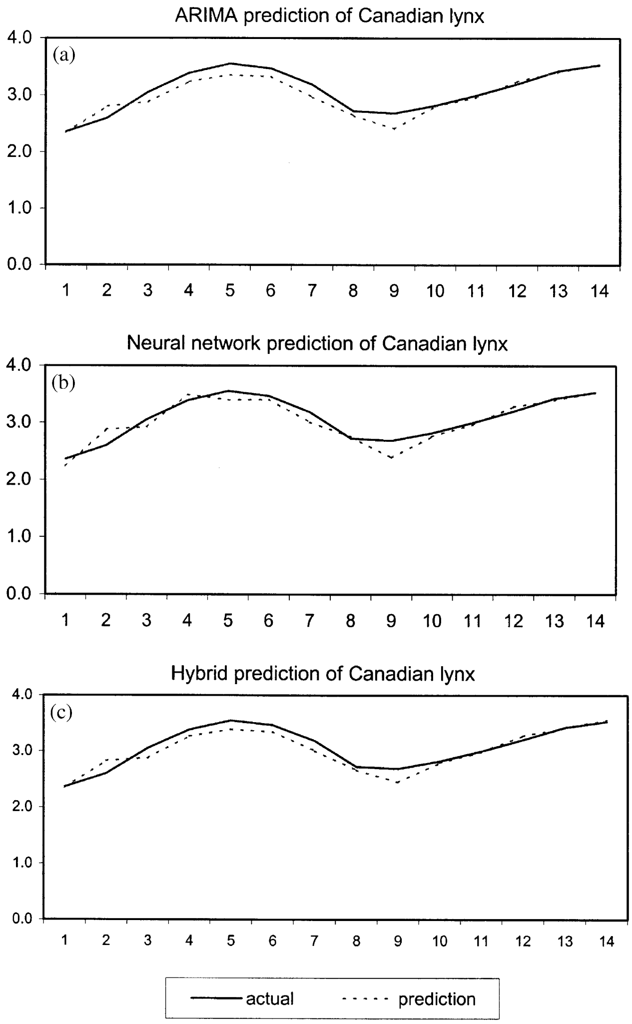

Fig. 5. (a) ARIMA prediction of Canadian lynx. (b) Neural network prediction of Canadian lynx. (c) Hybrid prediction of Canadian lynx.

Table 4
Exchange rate forecasting results[a]

	1 month		6 months		12 months	
	MSE	MAD	MSE	MAD	MSE	MAD
ARIMA	3.68493	0.005016	5.65747	0.0060447	4.52977	0.0053597
ANN	2.76375	0.004218	5.71096	0.0059458	4.52657	0.0052513
Hybrid	2.67259	0.004146	5.65507	0.0058823	4.35907	0.0051212

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

With the exchange rate data, the best linear ARIMA model is found to be the 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. This means that the most recent observation is the best guide for the next prediction. A neural network of $7 \times 6 \times 1$ is used to model the nonlinear patterns. Table 4 presents the test sample results with three time horizons of 1, 6 and 12 months. Results show that for short-term forecasting (1 month), both neural network and hybrid models are much better in accuracy than the random walk model. For longer time horizons, the ANN model gives a comparable performance to the ARIMA model. The hybrid model outperforms both ARIMA and ANN models consistently across three different time horizons and with both error measures although the improvement for longer horizons is not very impressive. The point-to-point comparison between actual and predicted values is given in Fig. 6.

5. Conclusions

Time series analysis and forecasting is an active research area over the last few decades. The accuracy of time series forecasting is fundamental to many decision processes and hence the research for improving the effectiveness of forecasting models has never stopped. With the efforts of Box and Jenkins [2], the ARIMA model has become one of the most popular methods in the forecasting research and practice. More recently, artificial neural networks have shown their promise in time series forecasting applications with their nonlinear modeling capability. Although both ARIMA and ANNs have the flexibility in modeling a variety of problems, none of them is the universal best model that can be used indiscriminately in every forecasting situation.

In this paper, we propose to take a combining approach to time series forecasting. The linear ARIMA model and the nonlinear ANN model are used jointly, aiming to capture different forms of relationship in the time series data. The hybrid model takes advantage of the unique strength of ARIMA and ANN in linear and nonlinear modeling. For complex problems that have both linear and nonlinear correlation structures, the combination method can be an effective way to improve forecasting performance. The empirical results with three real data sets clearly suggest that the hybrid model is able to outperform each component model used in isolation.

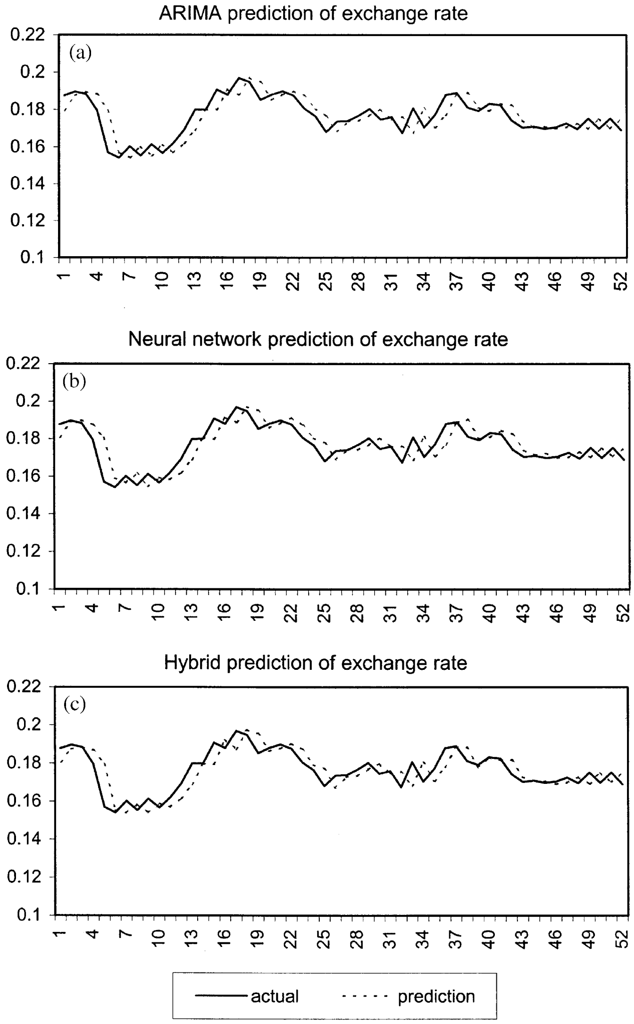

Fig. 6. (a) ARIMA prediction of exchange rate. (b) Neural network prediction of exchange rate. (c) Hybrid prediction of exchange rate.

Various combining methods have been proposed in the literature. However, most of them are designed to combine the similar methods. Linear models are often combined in the traditional forecasting literature. In the neural network literature, neural committee and neural network ensemble are often used to describe the combination of several neural networks. Although this type of combination has improved forecasting accuracy, we believe a more effective way of combining should be based on quite different models. Theoretical as well empirical evidences in the literature suggest that by using dissimilar models or models that disagree each other strongly, the hybrid model will have lower generalization variance or error [15,20,31]. Additionally, because of the possible unstable or changing patterns in the data, using the hybrid method can reduce the model uncertainty which typically occurred in statistical inference and time series forecasting [5]. Furthermore, by fitting the ARIMA model first to the data, the overfitting problem that is more strongly related to neural network models can be eased.

Acknowledgements

The author would like to thank two anonymous referees for their helpful comments.

References

[1] J.M. Bates, C.W.J. Granger, The combination of forecasts, Oper. Res. Q. 20 (1969) 451–468.
[2] G.E.P. Box, G. Jenkins, Time Series Analysis, Forecasting and Control, Holden-Day, San Francisco, CA, 1970.
[3] M.J. Campbell, A.M. Walker, A survey of statistical work on the MacKenzie River series of annual Canadian lynx trappings for the years 1821–1934, and a new analysis, J. R. Statist. Soc. Ser. A 140 (1977) 411–431.
[4] C. Chatfield, What is the 'best' method of forecasting? J. Appl. Statist. 15 (1988) 19–39.
[5] C. Chatfield, Model uncertainty and forecast accuracy, J. Forecasting 15 (1996) 495–508.
[6] R. Clemen, Combining forecasts: a review and annotated bibliography with discussion, Int. J. Forecasting 5 (1989) 559–608.
[7] M. Cottrell, B. Girard, Y. Girard, M. Mangeas, C. Muller, Neural modeling for time series: a statistical stepwise method for weight elimination, IEEE Trans. Neural Networks 6 (1995) 1355–1364.
[8] J.G. De Gooijer, K. Kumar, Some recent developments in non-linear time series modelling, testing, and forecasting, Int. J. Forecasting 8 (1992) 135–156.
[9] C. De Groot, D. Wurtz, Analysis of univariate time series with connectionist nets: a case study of two classical examples, Neurocomputing 3 (1991) 177–192.
[10] J.W. Denton, How good are neural networks for causal forecasting? J. Bus. Forecasting 14 (1995) 17–20.
[11] R.F. Engle, Autoregressive conditional heteroscedasticity with estimates of the variance of U.K. inflation, Econometrica 50 (1982) 987–1008.
[12] R. Fildes, S. Makridakis, The impact of empirical accuracy studies on time series analysis and forecasting, Int. Statist. Rev. 63 (1995) 289–308.
[13] I. Ginzburg, D. Horn, Combined neural networks for time series analysis, Adv. Neural Inf. Process. Systems 6 (1994) 224–231.
[14] C.W.J. Granger, A.P. Anderson, An Introduction to Bilinear Time Series Models, Vandenhoeck and Ruprecht, Göttingen, 1978.
[15] C.W.J. Granger, Combining forecasts—Twenty years later, J. Forecasting 8 (1989) 167–173.
[16] K.W. Hipel, A.I. McLeod, Time Series Modelling of Water Resources and Environmental Systems, Amsterdam, Elsevier, 1994.
[17] K. Hornik, M. Stinchicombe, H. White, Using multi-layer feedforward networks for universal approximation, Neural Networks 3 (1990) 551–560.
[18] M.S. Hung, J.W. Denton, Training neural networks with the GRG2 nonlinear optimizer, Eur. J. Oper. Res. 69 (1993) 83–91.
[19] G.M. Jenkins, Some practical aspects of forecasting in organisations, J. Forecasting 1 (1982) 3–21.
[20] A. Krogh, J. Vedelsby, Neural network ensembles, cross validation, and active learning, Adv. Neural Inf. Process. 7 (1995) 231–238.

[21] J.T. Luxhoj, J.O. Riis, B. Stensballe, A hybrid econometric-neural network modeling approach for sales forecasting, Int. J. Prod. Econ. 43 (1996) 175–192.
[22] S. Makridakis, Why combining works? Int. J. Forecasting 5 (1989) 601–603.
[23] S. Makridakis, A. Anderson, R. Carbone, R. Fildes, M. Hibdon, R. Lewandowski, J. Newton, E. Parzen, R. Winkler, The accuracy of extrapolation (time series) methods: results of a forecasting competition, J. Forecasting 1 (1982) 111–153.
[24] S. Makridakis, C. Chatfield, M. Hibon, M. Lawrence, T. Millers, K. Ord, L.F. Simmons, The M-2 competition: a real-life judgmentally based forecasting study, Int. J. Forecasting 9 (1993) 5–29.
[25] I.S. Markham, T.R. Rakes, The effect of sample size and variability of data on the comparative performance of artificial neural networks and regression, Comput. Oper. Res. 25 (1998) 251–263.
[26] E.D. McKenzie, General exponential smoothing and the equivalent ARMA process, J. Forecasting 3 (1984) 333–344.
[27] R.A. Meese, K. Rogoff, Empirical exchange rate models of the seventies: do they fit out of samples? J. Int. Econ. 14 (1983) 3–24.
[28] P. Newbold, C.W.J. Granger, Experience with forecasting univariate time series and the combination of forecasts (with discussion), J. R. Statist. Soc. Ser. A 137 (1974) 131–164.
[29] F.C. Palm, A. Zellner, To combine or not to combine? Issues of combining forecasts, J. Forecasting 11 (1992) 687–701.
[30] E. Pelikan, C. de Groot, D. Wurtz, Power consumption in West-Bohemia: improved forecasts with decorrelating connectionist networks, Neural Network World 2 (1992) 701–712.
[31] M.P. Perrone, L. Cooper, When networks disagree: ensemble method for hybrid neural networks, in: R.J. Mammone (Ed.), Neural Networks for Speech and Image Processing, Chapman & Hall, London, 1993, pp. 126–142.
[32] D.J. Reid, Combining three estimates of gross domestic product, Economica 35 (1968) 431–444.
[33] T. Subba Rao, M.M. Sabr, An Introduction to Bispectral Analysis and Bilinear Time Series Models, Lecture Notes in Statistics, Vol. 24, Springer-Verlag, New York, 1984.
[34] R. Sharda, R.B. Patil, Neural networks as forecasting experts: an empirical test, in: Proceedings of the International Joint Conference on Neural Networks, Washington, D.C., Vol. 2, 1990, pp. 491–494.
[35] V. Subramanian, M.S. Hung, A GRG2-based system for training neural networks: design and computational experience, ORSA J. Comput. 5 (1993) 386–394.
[36] Z. Tang, C. Almeida, P.A. Fishwick, Time series forecasting using neural networks vs Box–Jenkins methodology, Simulation 57 (1991) 303–310.
[37] Z. Tang, P.A. Fishwick, Feedforward neural nets as models for time series forecasting, ORSA J. Comput. 5 (1993) 374–385.
[38] H. Tong, Threshold Models in Non-Linear Time Series Analysis, Springer-Verlag, New York, 1983.
[39] D.K. Wedding II, K.J. Cios, Time series forecasting by combining RBF networks, certainty factors, and the Box–Jenkins model, Neurocomputing 10 (1996) 149–168.
[40] R. Winkler, Combining forecasts: a philosophical basis and some current issues, Int. J. Forecasting 5 (1989) 605–609.
[41] H. Wold, A Study in the Analysis of Stationary Time Series, Almgrist & Wiksell, Stockholm, 1938.
[42] G.U. Yule, 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 (1926) 1–64.
[43] G. Zhang, E.B. Patuwo, M.Y. Hu, Forecasting with artificial neural networks: the state of the art, Int. J. Forecasting 14 (1998) 35–62.
[44] G.P. Zhang, E.B. Patuwo, M.Y. Hu, A simulation study of artificial neural networks for nonlinear time-series forecasting, Comput. Oper. Res. 28 (2001) 381–396.
[45] M.M. Elkateb, K. Solaiman, Y. Al-Turki, A comparative study of medium-weather-dependent load forecasting using enhanced artificial/fuzzy neural network and statistical techniques, Neurocomputing 23 (1998) 3–13.
[46] T. Al-Saba, I. El-Amin, Artificial neural networks as applied to long-term demand forecasting, Artif. Intell. Eng. 13 (1999) 189–197.
[47] H.B. Hwang, Insights into neural-network forecasting time series corresponding to ARMA (p,q) structures, Omega 29 (2001) 273–289.

Guoqiang Peter Zhang is an Assistant Professor of Decision Science at Georgia State University. His research interests are statistical quality control, time series forecasting, neural networks, and supply chain management. His research has been published in journals such as Computers and Industrial Engineering, Computers and Operations Research, Decision Sciences, European Journal of Operational Research, IEEE Transactions on Systems, Man, and Cybernetics, International Journal of Forecasting, International Journal of Production Economics, Journal of Operational Research Society, and Omega, the International Journal of Management Science. He is on the editorial review board of Production and Operations Management.