AN EXTENDED FUZZY ARTIFICIAL NEURAL NETWORKS MODEL FOR TIME SERIES FORECASTING

M. KHASHEI, M. BIJARI AND S. R. HEJAZI

Abstract. Improving time series forecasting accuracy is an important yet often difficult task. Both theoretical and empirical findings have indicated that integration of several models is an effective way to improve predictive performance, especially when the models in combination are quite different. In this paper, a model of the hybrid artificial neural networks and fuzzy model is proposed for time series forecasting, using autoregressive integrated moving average models. In the proposed model, by first modeling the linear components, autoregressive integrated moving average models are combined with the these hybrid models to yield a more general and accurate forecasting model than the traditional hybrid artificial neural networks and fuzzy models. Empirical results for financial time series forecasting indicate that the proposed model exhibits effectively improved forecasting accuracy and hence is an appropriate forecasting tool for financial time series forecasting.

1. Introduction

Time series forecasting is an important area of forecasting. First, past observations on a variable are analyzed in order to develop a model describing the underlying relationship. This 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 dependent variable to other explanatory variables. The research for improving the effectiveness of forecasting models has been never stopped since accuracy of time series forecasting is fundamental to many decision processes. Several researchers have argued that predictive performance improves when we combine several models since the risk of an inappropriate model is reduced and hence the results are more accurate [18]. Typically, this is done when the underlying process cannot easily be determined; either we cannot identify the true data generating process or a single model may not be sufficient to identify all the characteristics of the time series [40].

Since the early work of Reid [35], and Bates and Granger [3], much effort has been devoted to develop and improve hybrid forecasting Models for time series. In their pioneering work on combined forecasts, Bates and Granger [3] showed that a

Received: April 2010; Revised: September 2010; Accepted: October 2010

Key words and phrases: Auto-regressive integrated moving average (ARIMA), Artificial neural networks (ANNs), Fuzzy regression, Fuzzy logic, Time series forecasting, Financial markets.

linear combination of forecasts have a smaller error variance than any of the individual methods. Since then, the studies on this topic have expanded dramatically. Makridakis et al. [29] claimed that using a hybrid model or combining several models has become a common practice for improving forecasting accuracy ever since the well-known M-competition in which a combination of forecasts from more than one model often leads to improved forecasting performance while Pelikan et al. [34], and Ginzburg and Horn [14] proposed combining several feedforward neural networks to improve time series forecasting accuracy. Clemen [10] provides a comprehensive review and annotated bibliography in this area.

Recently, several hybrid forecasting models have been developed by integrating different models together in order to improve prediction accuracy [30]. Chen and Wang [8] constructed a combination model incorporating a seasonal autoregressive integrated moving average (SARIMA) model and support vector machines (SVMs) for seasonal time series forecasting. Zhou and Hu [51] proposed a hybrid modeling and forecasting approach based on Grey and Box-Jenkins autoregressive moving average (ARMA) models. Armano et al. [2] presented a new hybrid approach, which integrates artificial neural networks (ANNs) with genetic algorithms (GAs) for stock market forecasting. Yu et al. [47] proposed a novel nonlinear ensemble forecasting model integrating generalized linear auto regression (GLAR) with artificial neural networks to obtain accurate predictions in the foreign exchange market. Tsaih et al. [41] presented a hybrid artificial intelligence (AI) approach, integrating the rule-based systems technique and artificial neural networks for S&P 500 stock index prediction. Tseng et al. [43] proposed a hybrid model called SARIMABP, combining the seasonal autoregressive integrated moving average models with the back-propagation (BP) neural networks to predict seasonal time series data.

Luxhoj et al. [28] presented a hybrid econometric and neural networks model for sales forecasting. Voort et al. [44] introduced a hybrid method called KARIMA using a Kohonen self-organizing maps (SOMs) and autoregressive integrated moving average models for short-term prediction. Chang et al. [6] developed a hybrid model by integrating a self organization map (SOM) neural network, genetic algorithms (GAs) and Fuzzy Rule Base (FRB) to forecast the future sales of a printed circuit board factory. Lin and Cobourn [26] combined the Takagi-Sugeno fuzzy system and a nonlinear regression (NLR) model for time series forecasting. Pai [33] proposed the hybrid ellipsoidal fuzzy system (HEFST) model to forecast regional electricity loads in Taiwan. Based on the basic concepts of artificial neural networks, Khashei et al. [25] proposed a new hybrid model for time series forecasting using fuzzy regression models to overcome the data limitation of artificial neural networks and yield more accurate results, especially in incomplete data situations.

In this paper, a model is proposed which, using autoregressive integrated moving average models extends the hybrid fuzzy artificial neural network model and has higher forecasting accuracy. In the proposed model, linear components of time series are first modeled by an autoregressive integrated moving average model. Then the set of the estimated values and residuals which respectively contain the linear and nonlinear relationships, are used as input values in the hybrid fuzzy artificial neural network model. In order to show its appropriateness and effectiveness, the

method is applied to financial markets forecasting problems and its performance is compared with the autoregressive integrated moving average (ARIMA), traditional hybrid fuzzy artificial neural network (FANNs), and adaptive network fuzzy inference systems (ANFIS) models, using three time series including the exchange rate (US $/Iran rial), gold price (gram/US $), and exchange rate (Euro/Iran rial). 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 are briefly reviewed. Hybrid fuzzy artificial neural network (FANNs) models are introduced in section 3 and the proposed model is formulated in section 4. In section 5, this model is applied to financial markets forecasting and its performance is compared with those of other fuzzy and non-fuzzy forecasting models. The final section of the paper contains some conclusions.

2. The Auto-regressive Integrated Moving Average (ARIMA) Models

For more than half a century, e autoregressive integrated moving average 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:

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

where, y_t and are the actual value and random error at time period t, respectively. $\phi(B) = 1 - \sum_{i=1}^{p} \phi_i B^i$ and $\theta(B) = 1 - \sum_{j=1}^{q} \theta_j B^j$ are polynomials in B of degree p and q, ϕ_i $(i = 1, 2, ..., p)$ and θ_j $(j = 1, 2, ..., q)$ are model parameters, $\bigtriangledown = (1 - B)$, B is the backward shift operator. p and q are often referred to as orders of the model and d is an integer, often referred to as the order of differencing. Random errors, a_t, are assumed to be independently and identically distributed with mean zero and constant variance σ^2 . Based on the earlier work of Yule [48] and Wold [45], Box and Jenkins [4] developed a practical approach to building autoregressive integrated moving average models, which had a fundamental impact on the time series analysis and forecasting applications.

The Box-Jenkins [4] 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 certain theoretical autocorrelation properties. By matching the empirical autocorrelation patterns with theory, it is often possible to identify one or more potential models for the given time series. Box and Jenkins [4] proposed to use the autocorrelation function (ACF) and the partial autocorrelation function (PACF) of the sample data as the basic tools for identifying the order of the autoregressive integrated moving average model. Other order selection methods have been proposed based on validity criteria, information-theoretic approaches (Akaike's information criterion (AIC) [36]) and the minimum description length (MDL) [23,19, 27]and recently approaches based on intelligent paradigms, such as neural networks [20], genetic algorithms [31, 32] or fuzzy systems [16] have been proposed to improve the accuracy of order selection of the

autoregressive integrated moving average models.

In the identification step, data transformation is often required to make the time series stationary. Stationarity is a necessary condition for 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 so 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, and, in particular, checking if the model assumptions about the errors, a_t, are satisfied.

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

3. The Fuzzy Artificial Neural Networks Models

The parameters of the artificial neural networks models, $(w_{i,j}(i = 0, 1, 2, ..., p\ j = 0, 1, 2, ..., q), w_j(j = 0, 1, 2, ..., q))$, (weights and biases) are crisp. In hybrid artificial neural networks and fuzzy regression models [20], fuzzy parameters in the form of triangular fuzzy numbers, $(\tilde{w}_{i,j}(i = 0, 1, 2, ..., p\ j = 0, 1, 2, ..., q), \tilde{w}_j(j = 0, 1, 2, ..., q))$, are used instead. In addition, this study adapts the methodology formulated by Ishibuchi and Tanaka [21] for the situation including a wide spread of the forecasted interval. A hybrid model is described using a fuzzy function with a fuzzy parameter:

$$\tilde{y}_t = f(\tilde{w}_0 + \sum_{j=1}^{q} \tilde{w}_j.g(\tilde{w}_{0,j} + \sum_{i=1}^{p} \tilde{w}_{i,j}.y_{t-i})), \tag{2}$$

where, y_t are observations and $\tilde{w}_{i,j}(i = 0, 1, 2, ..., p\ j = 0, 1, 2, ..., q), \tilde{w}_j(j = 0, 1, 2, ..., q)$, are fuzzy numbers. equation (2) is modified as follows:

$$\tilde{y}_t = f(\tilde{w}_0 + \sum_{j=1}^{q} \tilde{w}_j.\tilde{X}_{t,j}) = f(\sum_{j=0}^{q} \tilde{w}_j.\tilde{X}_{i,j}), \tag{3}$$

where, $\tilde{X}_{t,j} = g(\tilde{w}_{0,j} + \sum_{i=1}^{p} \tilde{w}_{i,j}.y_{t-1})$. Fuzzy parameters in the form of triangular fuzzy numbers $\tilde{w}_{i,j} = (a_{i,j}, b_{i,j}, c_{i,j})$ are used as follows:

$$\mu_{\tilde{w}_{i,j}}(w_{i,j}) = \begin{cases} \frac{1}{b_{i,j}-a_{i,j}}(w_{i,j} - a_{i,j}) & if\ a_{i,j} \le w_{i,j} \le b_{i,j}, \\ \\ \frac{1}{b_{i,j}-c_{i,j}}(w_{i,j} - c_{i,j}) & if\ b_{i,j} \le w_{i,j} \le c_{i,j}, \end{cases} \tag{4}$$

where $\mu_{\tilde{w}}(w_{i,j})$ is the membership function of the fuzzy set that represents the parameter. Applying the extension principle [1,17], it is clear that the membership of $\tilde{X}_{t,j} = g(\tilde{w}_{0,j} + \sum_{i=1}^{p} \tilde{w}_{i,j}.y_{t-1})$ in equation (3) is given as:

$$\mu_{\tilde{X}_{i,j}}(x_{i,j}) = \begin{cases} \frac{\left(X_{t,j} - g(\sum_{i=0}^{p} a_{i,j}.y_{t,i})\right)}{g(\sum_{i=0}^{p} b_{i,j}.y_{t,i}) - g(\sum_{i=0}^{p} a_{i,j}.y_{t,i})} \\ \qquad if\ g(\sum_{i=0}^{p} a_{i,j}.y_{t,i}) \leq X_{t,j} \leq g(\sum_{i=0}^{p} b_{i,j}.y_{t,i}), \\ \frac{\left(X_{t,j} - g(\sum_{i=0}^{p} c_{i,j}.y_{t,i})\right)}{g(\sum_{i=0}^{p} b_{i,j}.y_{t,i}) - g(\sum_{i=0}^{p} c_{i,j}.y_{t,i})} \\ \qquad if\ g(\sum_{i=0}^{p} b_{i,j}.y_{t,i}) \leq X_{t,j} \leq g(\sum_{i=0}^{p} c_{i,j}.y_{t,i}), \\ 0 \qquad otherwise \end{cases} \tag{5}$$

where, $y_{t,i} = 1\ (t = 1, 2, ..., k\ i = 0),\ y_{t,i} = y_{t-i}\ (t = 1, 2, ..., p)$. If the fuzzy numbers, $\tilde{X}_{t,j}$ are triangular with membership function equation (5) the membership function of $\tilde{w}_j$ will be as follows:

$$\mu_{\tilde{w}_j}(w_j) = \begin{cases} \frac{1}{e_j - d_j}(w_j - d_j) & if\ d_j \leq w_j \leq e_j, \\ \frac{1}{e_j - d_j}(w_j - f_j) & if\ e_j \leq w_j \leq f_j, \\ 0 & otherwise, \end{cases} \tag{6}$$

The membership function of $\tilde{y}_t = f(\tilde{w}_0 + \sum_{j=1}^{q} \tilde{w}_j.\tilde{X}_{t,j}) = f(\sum_{j=0}^{q} \tilde{w}_j.\tilde{X}_{t,j})$ is given by

$$\mu_{\tilde{Y}}(y_t) \cong \begin{cases} \frac{-B_1}{2A_1} + \left[\left(\frac{B_1}{2A_1}\right)^2 - \frac{C_1 - f^{-1}(y_t)}{A_1}\right]^{1/2} & if\ C_1 \leq f^{-1}(y_t) \leq C_3, \\ \frac{B_2}{2A_2} + \left[\left(\frac{B_2}{2A_2}\right)^2 - \frac{C_2 - f^{-1}(y_t)}{A_2}\right]^{1/2} & if\ C_3 \leq f^{-1}(y_t) \leq C_2, \\ 0 & otherwise \end{cases} \tag{7}$$

where,

$$A_1 = \sum_{j=0}^{q} (e_j - d_j).\left(g\left(\sum_{i=0}^{p} b_{i,j}.y_{t,i}\right) - g\left(\sum_{i=0}^{p} a_{i,j}.y_{t,i}\right)\right),$$

$$B_1 = \sum_{j=0}^{q} (d_j.\left(g\left(\sum_{i=0}^{p} b_{i,j}.y_{t,i}\right) - g\left(\sum_{i=0}^{p} a_{i,j}.y_{t,i}\right)\right) + g\left(\sum_{i=0}^{p} a_{i,j}.y_{t,i}\right).(e_j - d_j)),$$

$$A_2 = \sum_{j=0}^{q} (f_j - e_j).\left(g\left(\sum_{i=0}^{p} c_{i,j}.y_{t,i}\right) - g\left(\sum_{i=0}^{p} b_{i,j}.y_{t,i}\right)\right),$$

$$B_2 = \sum_{j=0}^{q} (f_j.\left(g\left(\sum_{i=0}^{p} c_{i,j}.y_{t,i}\right) - g\left(\sum_{i=0}^{p} b_{i,j}.y_{t,i}\right)\right) + g\left(\sum_{i=0}^{p} c_{i,j}.y_{t,i}\right).(f_j - e_j)),$$

$C_1 \sum_{j=0}^{q} (d_j.g(\sum_{i=0}^{p} a_{i,j}.y_{t,i}))$ $C_2 \sum_{j=0}^{q} (f_j.g(\sum_{i=0}^{p} c_{i,j}.y_{t,i}))$,

$C_3 \sum_{j=0}^{q} (e_j.g(\sum_{i=0}^{p} b_{i,j}.y_{t,i}))$,

Now considering a threshold level h for all membership function values of observations, the relevant nonlinear programming is

$$Min \ \sum_{t=1}^{k} \sum_{j=0}^{q} (f_j.g(\sum_{i=0}^{p} c_{i,j}.y_{t,i})) - (d_j.g(\sum_{i=0}^{p} a_{i,j}.y_{t,i}))$$

$$Subject\ to \begin{cases} \frac{-B_1}{2A_1} + \left[\left(\frac{B_1}{2A_1}\right)^2 - \frac{C_1 - f^{-1}(y_t)}{A_1}\right]^{1/2} \leq h \quad if\ C_1 \leq f^{-1}(y_t) \leq C_3, \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad for\ t = 1, 2, ..., k, \\ \frac{B_2}{2A_2} + \left[\left(\frac{B_2}{2A_2}\right)^2 - \frac{C_2 - f^{-1}(y_t)}{A_2}\right]^{1/2} \leq h \quad if\ C_3 \leq f^{-1}(y_t) \leq C_2, \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad for\ t = 1, 2, ..., k, \end{cases} \tag{8}$$

4. **Formulation of the Proposed Model**

Although artificial neural networks (ANNs) are flexible computing frameworks and universal approximators that can be applied to a wide range of forecasting problems with a high degree of accuracy, their performance in some situations, such as when dealing with linear problems, is inconsistent. Many papers in the literature are devoted to comparing artificial neural networks with linear forecasting methods [50]. Several studies have shown that artificial neural networks are significantly better than the conventional linear models and their forecast considerably and consistently more accurate. However, other studies have reported inconsistent results. Foster et al. find that artificial neural networks are significantly inferior to linear regression and a simple average of exponential smoothing methods [13]. Brace et al. [5] also find that the performance of artificial neural networks is not as good as many other statistical methods commonly used in load forecasting. Denton [11], with generated data for several different experimental conditions, shows that under ideal conditions, with all regression assumptions, there is little difference in predictability between artificial neural networks and linear regression, and that artificial neural networks perform better only under less than ideal conditions such as multicollinearity, model misspecification and the presence of outliers. Hann and Steurer [15] make comparisons between neural networks and linear models 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 [39] also compare the performance of linear models with neural networks. Their results show that linear autoregressive models outperform neural networks in some cases.

Many other researchers have also compared artificial neural networks and the corresponding traditional methods in their particular applications. Fishwick reports that the performance of artificial neural networks is worse than that of the

simple linear regression [12]. Tang et al. [37], and Tang and Fishwick [38] try to answer the question: under what conditions can neural networks forecasters perform better than the linear time series forecasting methods such as Box- Jenkins models? Some researchers believe that in the situations where artificial neural networks perform worse than linear statistical models, the reason may simply be that the data is linear without much disturbance and therefore it cannot be expected of artificial neural networks to do better [50]. However, since using artificial neural networks to model linear problems have yielded mixed results, it is not wise to apply neural networks blindly to any type of data.

Both artificial neural networks and autoregressive integrated moving average 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 autoregressive integrated moving average models to complex nonlinear problems or using artificial neural networks to model linear problems may be equally inappropriate, In order to overcome the limitations of each components model and improve the forecasting accuracy, using hybrid models or combining several models has become a common practice. Theoretical as well empirical evidence in the literature suggests that the hybrid model will have lower generalization variance or error if one uses dissimilar models, or models that disagree strongly. Moreover, because of possible unstable or changing patterns in the data, using the hybrid method can reduce the model uncertainty, which typically occurrs in statistical inference and time series forecasting [49].

In the proposed model, an autoregressive integrated moving average model is initially fitted to model the linear components (L_t) of the time series $\{y_t\}$. In this stage we estimate the actual values of the time series ($\hat{L_t}$) and model parameters as follows:

$$z_t = \phi_1 z_{t-1} + \phi_2 z_{t-2} + ... + \phi_p z_{t-p} + a_t - \theta_{p+1} a_{t-1} - \theta_{p+2} a_{t-2} - ... - \theta_{p+q} a_{t-q}, \quad (9)$$

where, $\theta_1, \theta_2, ..., \theta_q$ and $\phi_1, \phi_2, ..., \phi_p$ are the autoregressive integrated moving average parameters and $z_t = (1-B)^d (y_t - \mu)$ in which B is the backward shift operator, μ is the mean of the observations and d is an integer, often referred to as the order of differencing. Now, if the time series is composed of a linear autocorrelation structure and a nonlinear component ($y_t = N_t + L_t$), the autoregressive integrated moving average model cannot model the nonlinear patterns (N_t). Hence, the residuals of the autoregressive integrated moving average model will contain only the nonlinear patterns ($e_t = y_t - \hat{L}_t$). In the second phase, the set of the estimated values and residuals of the linear model (respectively containing the linear and nonlinear relationships) is input in the hybrid fuzzy artificial neural network model instead of the original time series dataset. Modeling residuals using hybrid fuzzy artificial neural network model helps to discover nonlinear relationships. With P input nodes, Q hidden nodes, and one output node with linear transfer function, the neural network model for the residuals is as follows:

$$e_t = f(e_{t-1}, ..., e_{t-p}) + \epsilon_t = w_0 + \sum_{j=1}^{Q} w_j g(w_{0j} + \sum_{i=1}^{p} w_{i,j} e_{t-i}) + \epsilon_t, \quad (10)$$

where, e_t are residuals of the linear model, $w_{i,j} (i = 0, 1, 2, ..., P \; j = 1, 2, ..., Q)$, $w_j (j = 0, 1, 2, ..., Q)$, are fuzzy numbers, f is a nonlinear function determined by

the neural network, and ϵ_t is the random error. If the forecast from (10) is denoted by N_t, the combined forecast is given by

$$\begin{aligned}\hat{z}_t &= \hat{L} + \hat{N}_t \\ &= (\phi_1 z_{t-1} + \phi_2 z_{t-2} + ... + \phi_p z_{t-p} - \theta_1 a_{t-1} - \theta_2 a_{t-2} - ... - \theta_q a_{t-q}) + \\ &+ (w_0 + \sum_{j=1}^{Q} w_j g(w_{0j} + \sum_{i=1}^{P} w_{i,j} e_{t-i})) \\ &= \sum_{i=1}^{p} \phi_i . z_{t-i} + \sum_{j=1}^{q} \theta_j . a_{t-j} + \sum_{k=0}^{Q} w_k . u_{t,k}. \end{aligned} \tag{11}$$

where, $u_{t,k} = g(w_{0k} + \sum_{l=1}^{P} w_{l,k} e_{t-1})$. If the model parameters $(\phi_i (i = 1, 2, ..., p), \theta_j (j = 1, 2, ..., q), w_{l,k} (l = 0, 1, 2, ..., P \; k = 1, 2, ..., Q), w_j (j = 0, 1, 2, ..., Q))$ are triangular fuzzy numbers $(\tilde{\phi}_i (i = 1, 2, ..., p), \tilde{\theta}_j (j = 1, 2, ..., q), \tilde{w}_{l,k} (l = 0, 1, 2, ..., P \; k = 1, 2, ..., Q), \tilde{w}_j (j = 0, 1, 2, ..., Q))$ the proposed fuzzy model is

$$\tilde{z}_t = \sum_{i=1}^{p} \tilde{\phi}_i . z_{t-i} + \sum_{j=1}^{q} \tilde{\theta}_j . a_{t-j} + \sum_{k=0}^{Q} \tilde{w}_k . \tilde{u}_{t,k}. \tag{12}$$

where, $\tilde{u}_{y,k} = g(\tilde{w}_{0k} + \sum_{l=1}^{P} \tilde{w}_{l,k} e_{t-1})$. Now, equation (12) is modified as follows:

$$\begin{aligned}\tilde{z}_t &= \sum_{i=l}^{p} \tilde{\beta}_i . \tilde{z}_{t-i} + \sum_{i=p+l}^{p+q} \tilde{\beta}_i . \tilde{a}_{t+p-i} + \sum_{i=p+q+l}^{p+q+Q+l} \tilde{\beta}_i . \tilde{u}_{t,i-(p+q)} \\ &= \sum_{i=1}^{p+q+Q+l} \tilde{\beta}_i . \tilde{x}_{t,i}. \end{aligned} \tag{13}$$

where, $\tilde{x}_{t,i} = g(\tilde{\alpha}_{0i} + \sum_{l=1}^{P} \tilde{\alpha}_{l,i} . e_{t-l})$. The fuzzy parameters are triangular fuzzy numbers $\tilde{\alpha}_{l,i} = (a_{l,i}, b_{l,i}, c_{l,i})$ with membership function

$$\mu_{\tilde{\alpha}_{i,j}}(\alpha_{i,j}) = \begin{cases} \frac{l}{b_{l,i} - a_{l,i}} (\alpha_{l,i} - a_{l,i}) & if \; a_{l,i} \le \alpha_{l,i} \le b_{l,i}, \\ \frac{l}{b_{l,i} - c_{l,i}} (\alpha_{l,i} - c_{l,i}) & if \; b_{l,i} \le \alpha_{l,i} \le c_{l,i}, \\ 0 & otherwise, \end{cases} \tag{14}$$

Applying the extension principle, it becomes clear that the membership of $\tilde{x}_{t,i} = g(\tilde{\alpha}_{0i} + \sum_{l=1}^{P} \tilde{\alpha}_{l,i} . e_{t-1})$ in equation (13) is given by

$$\mu_{\tilde{x}_{t,i}}(x_{t,i}) = \begin{cases} \frac{(x_{t,i} - g(\sum_{l=0}^{P} a_{l,i} . e_{t,l}))}{g(\sum_{l=0}^{P} b_{l,i} . e_{t,l}) - g(\sum_{l=0}^{P} a_{l,i} . e_{t,l})} \\ \qquad\qquad if \; g(\sum_{l=0}^{P} a_{l,i} . e_{t,l}) \le x_{t,i} \le g(\sum_{l=0}^{P} b_{l,i} . e_{t,l}), \\ \frac{(x_{t,i} - g(\sum_{l=0}^{P} c_{l,i} . e_{t,l}))}{g(\sum_{l=0}^{P} b_{l,i} . e_{t,l}) - g(\sum_{l=0}^{P} c_{l,i} . e_{t,l})} \\ \qquad\qquad if \; g(\sum_{l=0}^{P} b_{l,i} . e_{t,l}) \le x_{t,i} \le g(\sum_{l=0}^{P} c_{l,i} . e_{t,l}), \\ 0 \qquad\qquad otherwise \end{cases} \tag{15}$$

where, $e_{t,l} = l\ (t = 1, 2, ..., k\ l = 0),\ e_{t,l} = e_{t-l}\ (t = 1, 2, ..., k\ l = 1, 2, ..., p)$. Hence the membership function of the triangular fuzzy parameters ($\tilde{\beta}_i\ i = 1, 2, ..., p + q + Q + 1$) is

$$\mu_{\tilde{\beta}_i}(\beta_i) = \begin{cases} \frac{1}{r_i - d_i}(\beta_i - d_i) & if\ d_i \leq \beta_i \leq r_i, \\ \frac{1}{r_i - d_i}(\beta_i - f_i) & if\ r_i \leq \beta_i \leq f_i, \\ 0 & otherwise, \end{cases} \tag{16}$$

Also, the membership function of $\tilde{z}_t = \sum_{i=l}^{p+q+Q+l} \tilde{\beta}_i.\tilde{x}_{t,i}$ is given by

$$\mu_{\tilde{z}}(z_t) \cong \begin{cases} \frac{-B_1}{2A_1} + \left[\left(\frac{B_1}{2A_1}\right)^2 - \frac{C_1 - z_t}{A_1}\right]^{1/2} & if\ C_1 \leq z_t \leq C_3, \\ \frac{B_2}{2A_2} + \left[\left(\frac{B_2}{2A_2}\right)^2 - \frac{C_2 - z_t}{A_2}\right]^{1/2} & if\ C_3 \leq z_t \leq C_2, \\ 0 & otherwise \end{cases} \tag{17}$$

where,

$$A_1 = \sum_{j=0}^{p+q+Q}(r_j - d_j).\left(g\left(\sum_{i=0}^{P} b_{i,j}.e_{t,i}\right) - g\left(\sum_{i=0}^{P} a_{i,j}.e_{t,i}\right)\right),$$

$$B_1 = \sum_{j=0}^{p+q+Q}(d_j.(g(\sum_{i=0}^{P} b_{i,j}.e_{t,i}) - g(\sum_{i=0}^{P} a_{i,j}.e_{t,i})) + g(\sum_{i=0}^{P} a_{i,j}.e_{t,i}).(r_j - d_j)),$$

$$A_2 = \sum_{j=0}^{p+q+Q}(f_j - r_j).\left(g\left(\sum_{i=0}^{P} c_{i,j}.e_{t,i}\right) - g\left(\sum_{i=0}^{P} b_{i,j}.e_{t,i}\right)\right),$$

$$B_2 = \sum_{j=0}^{p+q+Q}(f_j.(g(\sum_{i=0}^{P} c_{i,j}.e_{t,i}) - g(\sum_{i=0}^{P} b_{i,j}.e_{t,i})) + g(\sum_{i=0}^{P} c_{i,j}.e_{t,i}).(f_j - r_j)),$$

$$C_1 = \sum_{j=0}^{p+q+Q}\left(d_j.g\left(\sum_{i=0}^{P} a_{i,j}.e_{t,i}\right)\right) \qquad C_2 = \sum_{j=0}^{p+q+Q}\left(f_j.g\left(\sum_{i=0}^{P} c_{i,j}.e_{t,i}\right)\right)$$

$$C_3 = \sum_{j=0}^{p+q+Q}\left(r_j.g\left(\sum_{i=0}^{P} b_{i,j}.e_{t,i}\right)\right)$$

Now considering a threshold level h for all membership function values of observations, we have the following nonlinear programming problem:

$$Min\ \sum_{t=1}^{k}\sum_{j=0}^{p+q+Q}\left(f_j.g\left(\sum_{i=0}^{P} c_{i,j}.e_{t,i}\right)\right) - \left(d_j.g\left(\sum_{i=0}^{P} a_{i,j}.e_{t,i}\right)\right)$$

$$Subject\ to \begin{cases} \frac{-B_1}{2A_1} + \left[\left(\frac{B_1}{2A_1}\right)^2 - \frac{C_1 - z_t}{A_1}\right]^{1/2} \leq h & if\ C_1 \leq z_t \leq C_3, \\ & for\ t = 1, 2, ..., k, \\ \frac{B_2}{2A_2} + \left[\left(\frac{B_2}{2A_2}\right)^2 - \frac{C_2 - z_t}{A_2}\right]^{1/2} \leq h & if\ C_3 \leq z_t \leq C_2, \\ & for\ t = 1, 2, ..., k, \end{cases} \tag{18}$$

As a special case and for simplicity and efficiency of in forecasting, the triangular fuzzy numbers are considered symmetric, and connected weights between input and hidden layer are considered to be crisp. Then the membership function of z_t transforms to

$$\mu_{\tilde{z}}(z_t) = \begin{cases} l - \frac{\left|z_t - \sum_{i=1}^{p} \alpha_i z_{t-i} + \sum_{i=p+l}^{p+q} \alpha_i a_{t+p-i} - \sum_{i=p+q+l}^{p+q+Q+l} \alpha_i u_{t,i-(p+q)}\right|}{\sum_{i=1}^{p} c_i|z_{t-i}| + \sum_{i=p+l}^{p+q} c_i|a_{t+p-i}| + \sum_{i=p+q+l}^{p+q+Q+l} c_i|u_{t,i-(p+q)}|} \\ \qquad\qquad for\ z_t \neq 0,\ a_t \neq 0,\ u_{t,i} \neq 0, \\ 0 \qquad\qquad otherwise \end{cases} \tag{19}$$

z_t represents the tth observation and h-level is the threshold value representing the degree to which the model should be satisfied by all the data points $y_1, y_2, ..., y_k$. The choice of h influences the widths of the fuzzy parameters:

$$\mu_{\tilde{z}}(z_t) \geq h \ for\ t = 1, 2, ..., k \tag{20}$$

The index t refers to the number of non-fuzzy data used for constructing the model. On the other hand, the fuzziness S included in the model is defined by:

$$S = \sum_{i=1}^{p}\sum_{t=1}^{k} c_i|\phi_i||z_{t-i}| + \sum_{i=p+l}^{p+q}\sum_{t=l}^{k} c_i|\rho_i - p||a_{t+p-i}| + \sum_{i=p+q+l}^{p+q+Q+l}\sum_{t=l}^{k} c_i|w_i||u_{t,i-(p+q+l)}|. \tag{21}$$

where, ρ_{i-p} is the autocorrelation coefficient of the time lag, $i-p$, ϕ_i is the partial autocorrelation coefficient of the time lag i, and w_i is the connection weight between an output neuron and the ith hidden neuron. Next, the problem of finding the parameters is formulated as a linear programming problem:

$$\begin{aligned} &Min\ S = \textstyle\sum_{i=l}^{p}\sum_{t=l}^{k} c_i|\phi_i||z_{t-i}| + \sum_{i=p+l}^{p+q}\sum_{t=l}^{k} c_i|\rho_i - p||a_{t+p-i}| + \\ &\quad \textstyle\sum_{i=p+q+l}^{p+q+Q+l}\sum_{t=l}^{k} c_i|w_i||u_{t,i-(p+q)}| \\ &S.T. \begin{cases} \sum_{i=l}^{p} \alpha_i z_{t-i} - \sum_{i=p+l}^{p+q} \alpha_i a_{t+p-i} + \sum_{i=p+q+1}^{p+q+Q+l} \alpha_i u_{t,i-(p+q)}\ t = 1, 2, ..., k \\ \quad +(1-h). \\ \quad \left(\sum_{i=1}^{p} c_i|z_{t-i} + \sum_{i=p+l}^{p+q} c_i|a_{t+p-i}| + \sum_{i=p+q+l}^{p+q+Q+l}\sum_{t=l}^{k} c_i|u_{t,i-(p+q)}|\right) \geq z_t \\ \sum_{i=l}^{p} \alpha_i z_{t-i} - \sum_{i=p+l}^{p+q} \alpha_i a_{t+p-i} + \sum_{i=p+q+1}^{p+q+Q+l} \alpha_i u_{t,i-(p+q)}\ t = 1, 2, ..., k \\ \quad -(1-h). \\ \quad \left(\sum_{i=1}^{p} c_i|z_{t-i}| + \sum_{i=p+l}^{p+q} c_i|a_{t+p-i}| + \sum_{i=p+q+l}^{p+q+Q+l}\sum_{t=l}^{k} c_i|u_{t,i-(p+q)}|\right) \geq z_t \\ c_i \geq 0 \qquad\qquad for\ i = 1, 2, ..., p+q+Q+1. \end{cases} \end{aligned} \tag{22}$$

The procedure is as follows:

Stage I (Phase I): *Linear modeling*: Fitting an autoregressive integrated moving average (ARIMA) model using the available information in the observations. The results of the phase I, are the optimum solution of the linear parameters, $\alpha^* = (\alpha_1^*, \alpha_2^*, ..., \alpha_{p+q}^*)$ and the residuals a_t (white noise), used as one of the input data sets in the next stage.

Stage I (Phase II): *Nonlinear modeling*: Training a neural network using the residual of autoregressive integrated moving average model. The results of phase II, are the optimum solution of the nonlinear parameters, $\alpha^* = (\alpha_{p+q+1}^*, \alpha_{p+q+2}^*, ..., \alpha_{p+q+Q}^*, \alpha_{p+q+Q+1}^*)$ and the output value of hidden neurons, used as one of the input data sets in the next stage.

Stage I (Phase III): *Combining*: Combining the results of phase I and phase II in order to model all relations in the time series data. The result of this phase is the final estimation of actual data.

Stage II: *Determining the minimal fuzziness*: Determining the minimal fuzziness using the same criterion as in the equation (18) and $\alpha^* = (\alpha_1^*, \alpha_2^*, ..., \alpha_{p+q+Q+1}^*)$. The number of constraint functions is the same as the number of observations. The proposed model is:

$$\begin{aligned} \tilde{z}_t &= \langle \alpha_1, c_1 \langle z_{t-1} + \langle \alpha_2, c_2 \langle z_{t-2} + ... + \langle \alpha_p, c_p \langle z_{t-p} - \langle \alpha_{p+1}, c_{p+1} \langle a_{t-1} \\ &- \langle \alpha_{p+2}, c_{p+2} \langle a_{t-2} - ... - \langle \alpha_{p+q}, c_{p+q} \langle a_{t-q} + \langle \alpha_{p+q+1}, c_{p+q+1} \langle u_{t,1} \\ &+ \langle \alpha_{p+q+2}, c_{p+q+2} \langle u_{t,2} + ... + \langle \alpha_{p+q+Q}, c_{p+q+Q} \langle u_{t,p+q+Q} \\ &+ \langle \alpha_{p+q+Q+1}, c_{p+q+Q+1} \langle u_{t,p+q+Q+1} \end{aligned} \tag{23}$$

where, $z_t = (1-B)^d (y_t - \mu)$, B is the backward shift operator , d is an integer, often referred to as order of differencing, μ is the mean of the observations, α_i is the center of the relevant fuzzy number, and c_i is the spread around its center.

Stage III: *Delete the outlying data*: According to Ishibuchi and Tanaka [21], when a model has outliers with wide spread, the data close to the upper and lower bounds of the model should be deleted and then the fuzzy regression model reformulated. In order to make the model include all possible conditions, we will say that c_j has a wide spread when the data set includes a significant difference or an outlying case.

5. Application of the Proposed Model to Financial Markets Forecasting

In order to demonstrate the appropriateness and effectiveness of the proposed model, we consider forecasting the gold price (*Gram/US* \$), exchange rate (*US* \$*/Iran rial*), and exchange rate (*Euro/Iran rial*) usinf time series.

5.1. Gold Price (Gram/US$) Forecasting.

The information used in this investigation consists of 40 daily observations (Figure 1) of gold price (Gram/US \$) from *26 November 2005* to *18 January 2006*. The first 35 observations are used to formulate the model and the last five to evaluate the model performance.

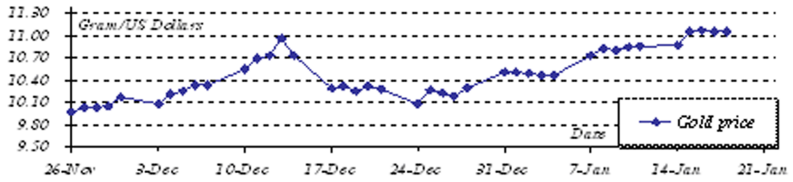


FIGURE 1. Gold Price (Gram/US $) from *26 November 2005* to *18 January 2006*. Reference: Central Bank of Iran (*CBI*)

Stage I (Phase I): *Linear modeling*: Using *Eviews* package software, the best-fitted model is *ARIMA (2, 1, 0)* ad we have

$$\hat{y}_t = 10.76 + 1.039y_{t-1} - 0.124y_{t-2}. \tag{24}$$

Stage I (Phase II): *Nonlinear modeling*: In order to obtain the optimum network architecture, several network architectures are evaluated and the performance of the ANNs are compared using the concepts of artificial neural networks design [24] and the constructive algorithm in *MATLAB7* package software. The best fitted network is selected, and the architecture which presents the best forecasting accuracy with the test data, is composed of three inputs, three hidden and one output neurons (in abbreviated form, $N^{(4-3-1)}$). The structure of this network is shown in Figure 2 , its weights and biases are given in Table 1 and the performance measures are given in Table 2.

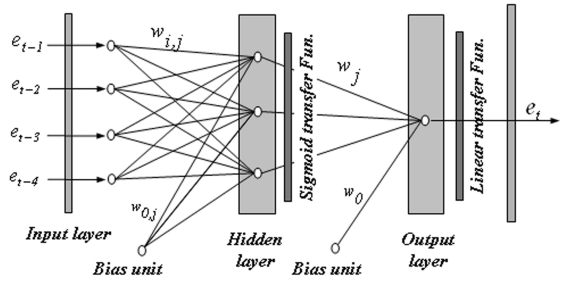


FIGURE 2. Structure of the Best Fitted Network (Gold Price Case), $N^{(4-3-1)}$

Stage II: *Determining the minimal fuzziness*: Setting $(\alpha_0^*, \alpha_1^*, \alpha_2^*) = (10.76, 1.039, 0.124)$ and $(\alpha_3^*, \alpha_4^*, \alpha_5^*, \alpha_6^*) = (2.0626, -0.03325, -1.0947, 3.0331)$, as in the previous example, the fuzzy parameters are obtained using equation (22) (with $h = 0$). These results are plotted in Figure 3.

Input Weights				Hidden Weights	Biases	
$W_{i,1}$	$W_{i,2}$	$W_{i,3}$	$W_{i,3}$		W_{0j}	W_0
5.88 8.3397	4.6177 -12.4315	1.2991 13.2273	-6.0498 2.7559	-0.93325 -1.0947	-7.0249 -12.8186	-2.0626

TABLE 1. Weights and Biases of the Designed Network (Gold Price Case)

Train		Test					
MSE	MAE	MSE	SSE	RMSE	ME	MAPE	MAE
0.0058	**0.0530**	**0.0043**	0.0217	0.0659	0.0403	0.5859	**0.0646**

TABLE 2. Performance Measures of the Designed Network (Gold Price Case)

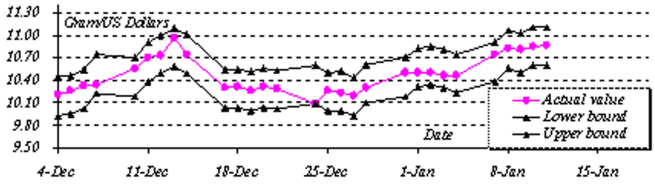


FIGURE 3. Results Obtained from the Proposed Model (Gold Price Case)

The hybrid method provides the possible intervals. From Figure 3, we conclude that the actual values are located in fuzzy intervals but the thread of these intervals is not satisfactory, especially when the macro-economic environment is smooth. Therefore, the Ishibuchi and Tanaka method [21] is used to resolve the problem and to provide a narrower interval for decision making.

Stage III: *Delete the outlying data*: It is clear that the observation of *24-December* is located at the lower bound, so the LP constrained equation that is produced by this observation is deleted and there is a return to stage II. The new results are shown in Figure 4. The future value of the gold price (Gram/US $) of the last five transaction days is forecasted using the revised model proposed and the results are shown in Table 3.

In time series analysis, it is assumed that the residuals $(y_t - \hat{y}_t)$ are white noise. If this condition is not satisfied, the forecasts will not be reliable. Unfortunately,neural networks time series analysis does not check the residuals for this property. Therefore, the residuals of the proposed model are plotted in order to check their behavior. The results (Figure 5) indicate that the residuals of the proposed model are identically distributed around zero, so they are white noise and their estimates are reliable.

Date	Actual	Lower bound	Upper bound
14- January	10.87	10.81	11.10
15- January	11.06	10.86	11.15
16- January	11.08	10.99	11.29
17- January	11.06	10.94	11.23
18- January	11.06	11.00	11.29

TABLE 3. Results of the Proposed Model for Test Data (Gold Price Case)

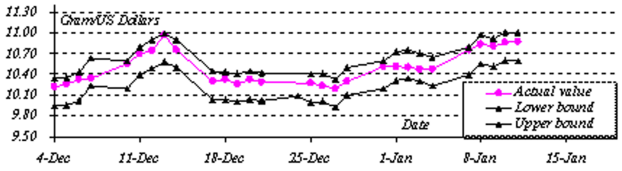


FIGURE 4. Results Obtained from the Proposed Model (After Deleting)

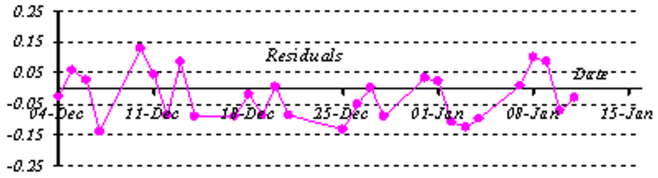


FIGURE 5. Residuals of the Proposed Model (Gold Price Case)

5.2. **Exchange Rate (US Dollar/ Iran Rial) Forecasting.** The information used in this investigation consists of 42 daily observations of the exchange rate of United States dollars against Iran rial from *5 Nov to 16 Dec, 2006* (Figure 6). The first 35 observations (five weeks) are used to formulate the model and the last seven (last week) to evaluate its performance.

Stage I (Phase I): *Linear modeling*: As in the previous example, the best model is $ARIMA(2,1,0)$ as given below:

$$\hat{y}_t = 9060.5 + 0.607y_{t-1} + 0.421y_{t-2}. \tag{25}$$

Stage I (Phase II): *Nonlinear modeling*: Again, the best network selected is composed of two inputs, three hidden and one output neurons (in abbreviated form, $N^{(3-3-1)}$). The structure of the this network is shown in Figure 7, the weights and biases of are given in Table 4 and the performance measures of are given in Table 5.

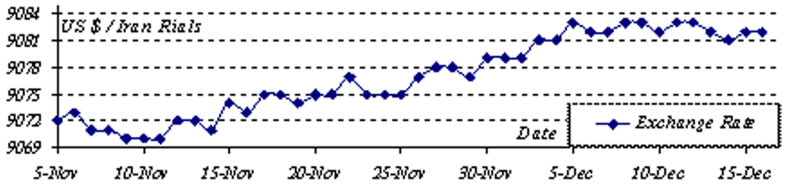


FIGURE 6. Exchange Rate (US Dollars/ Iran Rial) from 5 Nov to 16 Dec 2006. Reference: Central Bank of Iran (CBI)

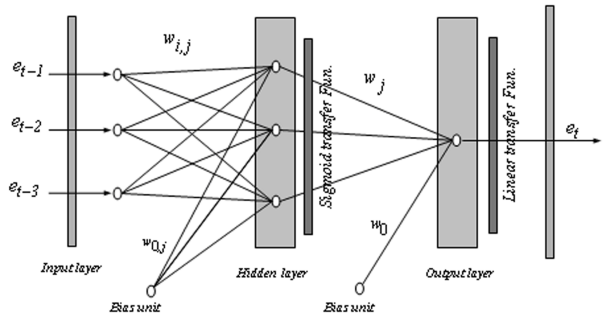


FIGURE 7. Structure of the Best Fitted Network (Exchange Rate Case), N(3-3-1)

Input Weights			*Hidden*	*Biases*	
$W_{i,1}$	$W_{i,2}$	$W_{i,3}$	*Weights*	W_{0j}	W_0
-3.3947	1.9112	11.9393	3.3696	-2.7097	
-2.3734	-0.31743	-28.5843	6.2053	2.3013	-4.0598
6.8024	-3.6088	-27.0192	-1.1491	28.9588	

TABLE 4. Weights and Biases of the Designed Network (Exchange Rate Case)

Train		*Test*					
MSE	*MAE*	*MSE*	*SSE*	*RMSE*	*ME*	*MAPE*	*MAE*
0.3875	**0.4421**	**0.4782**	3.3475	0.6915	0.0665	0.0068	**0.6138**

TABLE 5. Performance Measures of the Designed Network (Exchange Rate Case)

Stage II: *Determining the minimal fuzziness*: Setting $(\alpha_0^*, \alpha_1^*, \alpha_2^*) = (9060.05, 0.607, 0.421)$ and $(\alpha_3^*, \alpha_4^*, \alpha_5^*, \alpha_6^*) = (3.37, 6.205, -1.149, -4.060)$, the fuzzy parameters are obtained using equation (22) (with $h = 0$). The results are plotted in Figure 8.

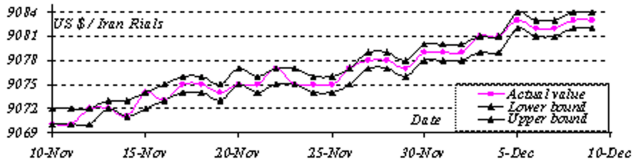


FIGURE 8. Results Obtained from the Proposed Model (Exchange Rate Case)

Stage III: *Delete the outlying data*: It is clear that the observation of *22-November* is located at the upper bound, so the LP constrained equation produced by this observation is deleted and there is a return to stage II. The new results are shown in Figure 9. The residuals of the model are plotted in Figure 10. The results indicate that these residuals are also identically distributed around zero, so they are white noise and their estimates are reliable.

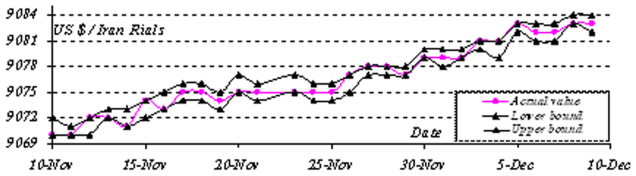


FIGURE 9. Results Obtained from the Proposed Model (After Deleting)

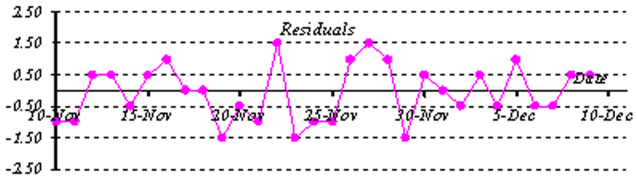


FIGURE 10. Residuals of the Proposed Model (Exchange Rate Case)

Using the revised hybrid model, the future value of the exchange rate (US $/ Iran rial) of the last seven transaction days is forecasted and the results are shown in Table 6. Although the proposed model is specifically proposed for forecasting situations with scant historical data available, the performance of the proposed model can be improved with larger data sets, as is the case with other quantitative forecasting models.

Date	Actual	Lower bound	Upper bound
1- December	9082	9081	9083
2- December	9083	9082	9083
3- December	9083	9082	9083
4- December	9082	9082	9084
5- December	9081	9080	9081
6- December	9082	9080	9082
7- December	9082	9082	9084

TABLE 6. Results of the Proposed Model for Test Data (Exchange Rate Case)

5.3. **Comparison with Other Forecasting Models.** In this section, the predictive capability of the proposed model is compared with the autoregressive integrated moving average (ARIMA), the hybrid artificial neural networks and fuzzy (FANNs) [24], and the adaptive neuro-fuzzy inference systems (ANFIS) [22] models, using three time series (the United States dollar against the Iran rial exchange rate, the gold price (Gram/US dollar), and the Euro against the Iran rial exchange rate). Other fuzzy and nonfuzzy forecasting models such as Chen's fuzzy time series (first-order) [7], Chen's fuzzy time series (high-order) [9], Yu's fuzzy time series [46], fuzzy autoregressive integrated moving average (FARIMA) [42], and artificial neural networks (ANNs) have also been considered for comparison with the forecasting power of the proposed model in point and interval forecasting cases. To measure forecasting performance in the interval and point estimation cases, the width of the forecasted interval, and MAE (Mean Absolute Error), and MSE (Mean Squared Error) are respectively employed as performance indicators, and respectively computed from the following equations.

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|e_i| \tag{26}$$

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(e_i)^2 \tag{27}$$

Based on the results obtained from these cases (Table 7), the predictive capabilities of the proposed model are rather encouraging and the possible interval suggested by the proposed model is narrower than the possible interval of the hybrid artificial neural networks and fuzzy model. The width of the forecasted interval in the proposed model is 0.29 dollar in the gold price (Gram/US dollar), 1.6 and 11.4 rial in the exchange rate (US dollar and Euro/Iran rial) forecasting cases, indicating a 9.4%, 36.0%, and 15.6% improvement upon the possible interval of the hybrid artificial neural networks and fuzzy model, respectively.

Similarly, the width of the forecasted interval is narrower than those obtained by ARIMA (95% Confidence Interval), and fuzzy ARIMA models.

In addition, according to the numerical results (Table 8 and 9), the MAE and MSE of the proposed model are lower than the hybrid artificial neural networks and fuzzy model, except for the MAE in an exchange rate (US dollar/Iran rial) case. For example in terms of MSE, the percentage improvements of the proposed model over the hybrid model, are 66.7% , 8.8%, and 8.4%, in the gold price, the exchange rates of the United States dollar and the Euro against the Iran rial cases, respectively.

Model	Cases	Forecasted interval width	Related Performance		
			ARIMA (95% Confidence)	Fuzzy ARIMA	Hybrid ANNs and fuzzy
ARIMA (95% Con. Interval)	Gold price	1.22	0	-	-
Fuzzy ARIMA		0.57	53.3%	0	-
Hybrid ANNs and fuzzy		0.32	73.8%	43.9%	0
Proposed Model		0.29	76.2%	49.1%	9.4%
ARIMA (95% Conf. Interval)	US dollar /Iran rial	16.2	0	-	-
Fuzzy ARIMA		4.2	74.1%	0	-
Hybrid ANNs and fuzzy		2.5	84.6%	40.5%	0
Proposed Model		1.6	90.1%	61.9%	36.0%
ARIMA (95% Conf. Interval)	Euro / Iran rial	66.9	0	-	-
Fuzzy ARIMA		23.7	64.6%	0	-
Hybrid ANNs and fuzzy		13.5	79.8%	43.0%	0
Proposed Model		11.4	82.9%	51.9%	15.6%

TABLE 7. Comparison of Forecasted Interval Widths by the Proposed Model Compared to Other Models (Interval Estimation)

Similarity, the MAE and MSE of the proposed model are lower than those of Chen's fuzzy time series (first-order and second-order), Yu's fuzzy time series, adaptive neuro-fuzzy inference system (ANFIS), autoregressive integrated moving average (ARIMA), and artificial neural networks (ANNs) in all cases, except for the MSE of the Yu's model for exchange rate (Euro/Iran rial) forecasting.

6. **Conclusions**

Time series forecasting is an active research area with applications in a variety of fields. Despite the numerous time series models available, the accuracy of time series forecasting is fundamental to many decision processes and hence, research into ways of improving the effectiveness of forecasting models has never been given up. Combining several models or using hybrid models can be an effective way to improve forecasting performance. Theoretical as well empirical evidence in the literature suggests that by using dissimilar models or models that disagree strongly, a hybrid model will have lower generalization variance or error. Moreover, because of the possible unstable or changing patterns in the data, using the hybrid method can reduce model uncertainty, which typically occurrs in statistical inference and time series forecasting.

In this paper, a model is proposed which, using autoregressive integrated moving average (ARIMA) models, extends hybrid artificial neural networks and fuzzy (FANNs) to yield more accurate results. In the proposed model, linear components of time series are first modeled by an autoregressive integrated moving average model and then , instead of the original time series data, the set of the estimated values and residuals, respectively containing only the linear and the nonlinear relationships, are applied as input values in the hybrid artificial neural networks and the fuzzy model. Experimental results of financial markets forecasting indicate that the proposed model is generally better than the hybrid artificial neural networks and fuzzy models and also better than the fuzzy and nonfuzzy models surveyed in this paper.

Model	Gold price (Gram /US dollar)		Exchange rate (US dollar / Iran rial)		Exchange rate (Euro / Iran rial)	
	MAE	MSE	MAE	MSE	MAE	MSE
Auto-Regressive Integrated Moving Average	0.105	0.017	0.924	1.24	56.44	3443
Chen's fuzzy time series (first-order)	0.297	0.116	0.750	0.777	48.60	3324
Chen's fuzzy time series (second-order)	0.292	0.111	0.750	0.777	48.60	3324
Yu's fuzzy time series	0.297	0.116	0.750	0.777	40.27	1783
Artificial Neural Networks (ANNs)	0.170	0.034	0.692	0.686	43.1	2173
Adaptive Neuro-Fuzzy Inference Systems	0.081	0.013	0.625	0.547	40.89	1979
Hybrid ANNs and fuzzy	0.097	0.012	0.583	0.524	39.80	2075
Proposed Model	0.065	0.004	0.614	0.478	33.49	1901

TABLE 8. Comparison of the Performance of the Proposed Model with Those of Other Forecasting Models (Point Estimation)

Model	Gold price (Gram /US dollar)		Exchange rate (US dollars / Iran rial)		Exchange rate (Euro / Iran rial)	
	MAE	MSE	MAE	MSE	MAE	MSE
Auto-Regressive Integrated Moving Average	38.1%	76.5%	33.5%	61.5%	40.7%	44.8%
Chen's fuzzy time series (first-order)	78.1%	96.6%	18.1%	38.5%	31.1%	42.8%
Chen's fuzzy time series (second-order)	77.7%	96.4%	18.1%	38.5%	31.1%	42.8%
Yu's fuzzy time series	78.1%	96.6%	18.1%	38.5%	16.8%	-6.6%
Artificial Neural Networks (ANNs)	61.8%	88.2%	11.3%	30.3%	22.3%	12.5%
Adaptive Neuro-Fuzzy Inference Systems	19.8%	69.2%	1.8%	12.6%	18.1%	3.9%
Hybrid ANNs and fuzzy	33.0%	66.7%	-5.3%	8.8%	15.9%	8.4%

TABLE 9. Improvement Percentage of the Proposed Model in Comparison with Those of Other Forecasting Models in Point Estimation

This evidence indicates that the proposed model can be an effective way to improve forecasting accuracy; therefore, it can be used as an alternative forecasting tool for financial markets forecasting.

Acknowledgements. The authors wish to express their gratitude to anonymous referees, Dr. G. A. Raissi Ardali, and Dr. A. K Tavakoli, Industrial Engineering Department, Isfahan University of Technology, for their insightful and constructive comments, which helped to improve the paper greatly.

REFERENCES

[1] A. R. Arabpour and M. Tata, *Estimating the parameters of a fuzzy linear regression model*, Iranian Journal of Fuzzy Systems, **5** (2008), 1-19.

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

[3] J. M. Bates and W. J. Granger, *The combination of forecasts*, Operation Research, **20** (1969), 451-468.

[4] P. Box and G. M. Jenkins, *Time series analysis: forecasting and control*, Holden-day Inc, San Francisco, CA, 1976.
[5] M. C. Brace, J. Schmidt and M. Hadlin, *Comparison of the forecasting accuracy of neural networks with other established techniques*, In: Proceedings of the First Forum on Application for weight elimination, IEEE Transactions on Neural Networks of Neural Networks to Power Systems, Seattle, WA (1991), 31-35.
[6] P. Chang, C. Liu and Y. Wang, *A hybrid model by clustering and evolving fuzzy rules for sales decision supports in printed circuit board industry*, Decision Support Systems, **42** (2006), 1254-1269.
[7] S. M. Chen, *Forecasting enrollments based on fuzzy time series*, Fuzzy Sets and Systems, **81(3)** (1996), 311–319, 1996.
[8] K. Y. Chen and C. H. Wang, *A hybrid SARIMA and support vector machines in forecasting the production values of the machinery industry in Taiwan*, Expert Systems with Applications, **32** (2007), 254-264.
[9] S. M. Chen and N. Y. Chung, *Forecasting enrollments using high-order fuzzy time series and genetic algorithms*, International J. Intell. Syst., **21** (2006), 485-501.
[10] R. Clemen, *Combining forecasts: a review and annotated bibliography with discussion*, International Journal of Forecasting, **5** (1989), 559-608.
[11] J. W. Denton, *How good are neural networks for causal forecasting?*, The Journal of Business Forecasting, **14(2)** (1995), 17-20.
[12] P. A. Fishwick, *Neural network models in simulation: a comparison with traditional modeling approaches*, In: Proceedings of Winter Simulation Conference, Washington, D. C., (1989), 702-710.
[13] W. R. Foster, F. Collopy and L. H. Ungar, *Neural network forecasting of short, noisy time series*, Computers and Chemical Engineering, **16(4)** (1992), 293-297.
[14] I. Ginzburg and D. Horn, *Combined neural networks for time series analysis*, Neural Information Processing Systems, **6** (1994), 224-231.
[15] T. H. Hann and E. Steurer, *Much ado about nothing? exchange rate forecasting: neural networks vs. linear models using monthly and weekly data*, Neurocomputing, **10** (1996), 323-339.
[16] M. Haseyama and H. Kitajima, *An ARMA order selection method with fuzzy reasoning*, Signal Process, **81** (2001), 1331-1335.
[17] H. Hassanpour, H. R. Maleki and M. A. Yaghoobi, *A note on evaluation of fuzzy linear regression models by comparing membership functions*, Iranian Journal of Fuzzy Systems, **6** (2009), 1-6.
[18] M. Hibon and T. Evgeniou, *To combine or not to combine: selecting among forecasts and their combinations*, International Journal of Forecasting, **21** (2005), 15-24.
[19] C. M. Hurvich and C. L. Tsai, *Regression and time series model selection in small samples*, Biometrica, **76(2)** (1989), 297-307.
[20] H. B. Hwang, *Insights into neural-network forecasting time series corresponding to ARMA(p; q) structures*, Omega, **29** (2001), 273-289.
[21] H. Ishibuchi and H. Tanaka, *Interval regression analysis based on mixed 0-1 integer programming problem*, J. Japan Soc. Ind. Eng, **40(5)** (1988), 312-319.
[22] J. S. R. Jang, *ANFIS: adaptive-network-based fuzzy inference system*, IEEE Trans Syst, Man, Cybernet, **23** (1993), 665-85.
[23] R. H. Jones, *Fitting autoregressions*, J. Amer. Statist. Assoc., **70(351)** (1975), 590-592.
[24] M. Khashei, *Forecasting the Isfahan Steel Company production price in Tehran Metals Exchange using Artificial Neural Networks (ANNs)*, Master of Science Thesis, Isfahan University of Technology, 2005.
[25] M. Khashei, S. R. Hejazi and M. Bijari, *A new hybrid artificial neural networks and fuzzy regression model for time series forecasting*, Fuzzy Sets and Systems, **159** (2008), 769-786.
[26] Y. Lin and W. G. Cobourn, *Fuzzy system models combined with nonlinear regression for daily ground-level ozone predictions*, Atmospheric Environment, **41** (2007), 3502-3513.

[27] L. Ljung, *System Identification Theory for the User, Prentice-Hall*, Englewood Cliffs, NJ, 1987.
[28] J. T. Luxhoj, J. O. Riis and B. Stensballe, *A hybrid econometric-neural network modeling approach for sales forecasting*, Int. J. Prod. Econ., **43** (1996), 175-192.
[29] S. Makridakis, A. Anderson, R. Carbone, R. Fildes. M. Hibdon, R. Lewandowski, J. Newton, E. Parzen and R. Winkler, *The accuracy of extrapolation (time series) methods: results of a forecasting competition*, Journal of Forecasting, **1** (1982), 111-53.
[30] E. Mehdizadeh, S. Sadi-nezhad and R. Tavakkoli-moghaddam, *Optimization of fuzzy clustering criteria by a hybrid pso and fuzzy c-means clustering algorithm*, Iranian Journal of Fuzzy Systems, **5** (2008), 1-14
[31] T. Minerva and I. Poli, *Building ARMA models with genetic algorithms*, In: Lecture Notes in Computer Science, **2037** (2001), 335-342.
[32] C. Ong, J. J. Huang and G. H. Tzeng, *Model identification of ARIMA family using genetic algorithms*, Appl. Math. Comput., **164(3)** (2005), 885-912.
[33] P. F. Pai and C. S. Lin, *A hybrid ARIMA and support vector machines model in stock price forecasting*, Omega, **33** (2005), 497-505.
[34] E. Pelikan, C. De Groot and D. Wurtz, *Power consumption in West-Bohemia: improved forecasts with decorrelating connectionist networks*, Neural Network, **2** (1992), 701-712.
[35] M. J. Reid, *Combining three estimates of gross domestic product*, Economica, **35** (1968), 431-444.
[36] R. Shibata, *Selection of the order of an autoregressive model by Akaike's information criterion*, Biometrika, **AC-63(1)** (1976), 117-126.
[37] Z. Tang, C. Almeida and P. A. Fishwick, *Time series forecasting using neural networks us,* Box-Jenkins Methodology Simulation, **57(5)** (1991), 303-310.
[38] Z. Tang and P. A. Fishwick, *Feedforward neural nets as models for time series forecasting*, ORSA Journal on Computing, **5(4)** (1993), 374-385.
[39] T. Taskaya and M. C. Casey, *A comparative study of autoregressive neural network hybrids*, Neural Networks, **18** (2005), 781-789.
[40] N. Terui and H. van Dijk, *Combined forecasts from linear and nonlinear time series models*, International Journal of Forecasting, **18** (2002), 421-438.
[41] R. Tsaih, Y. Hsu and C. C. Lai, *Forecasting S&P 500 stock index futures with a hybrid AI system*, Decision Support Systems, **23** (1998), 161-174.
[42] F. M. Tseng, G. H. Tzeng, H. C. Yu and B. J. C. Yuan, *Fuzzy ARIMA model for forecasting the foreign exchange market*, Fuzzy Sets and Systems, **118** (2001), 9-19.
[43] F. M. Tseng, H. C. Yu and G. H. Tzeng, *Combining neural network model with seasonal time series ARIMA model*, Technological Forecasting & Social Change, **69** (2002), 71-87.
[44] M. V. D. Voort and M. Dougherty and S. Watson, *Combining Kohonen maps with ARIMA time series models to forecast traffic flow*, Transportation Research Part C: Emerging Technologies, **4** (1996), 307-318.
[45] H. Wold, *A Study in the analysis of stationary time series*, Almgrist & Wiksell, Stockholm, 1938.
[46] H. K. Yu, *Weighted fuzzy time-series models for TAIEX forecasting*, Physica A, **349** (2004), 609-624.
[47] L. Yu, S. Wang and K. K. Lai, *A novel nonlinear ensemble forecasting model incorporating GLAR and ANN for foreign exchange rates*, Computers and Operations Research, **32** (2005), 2523-2541.
[48] G. 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.
[49] G. P. Zhang, *Time series forecasting using a hybrid ARIMA and neural network model*, Neurocomputing, **50** (2003), 159-175.
[50] G. Zhang, B. E. Patuwo and M. Y. Hu, *Forecasting with artificial neural networks: the state of the art*, International Journal of Forecasting, **14** (1998), 35-62.
[51] Z. J. Zhou and C. H. Hu, *An effective hybrid approach based on grey and ARMA for forecasting gyro drift, Chaos*, Solitons and Fractals, **35** (2008), 525-529.

Mehdi Khashe*, Industrial Engineering Department, Isfahan University of Technology, Isfahan, Iran
E-mail address: Khashei@in.iut.ac.ir

Mehdi Bijari, Industrial Engineering Department, Isfahan University of Technology, Isfahan, Iran
E-mail address: Bijari@cc.iut.ac.ir

Seyed Reza Hejazi, Industrial Engineering Department, Isfahan University of Technology, Isfahan, Iran
E-mail address: Rehejazi@cc.iut.ac.ir

*Corresponding author