*Journal of Industrial and Systems Engineering*
Vol. 10, No. 4, pp. 96-109
Autumn (November) 2017

# Comparison of Autoregressive Integrated Moving Average (ARIMA) model and Adaptive Neuro-Fuzzy Inference System (ANFIS) model (Case study: forecasting the gold price)

**Kazem Noghondarian**[1], **Emran Mohammadi**[1*], **Ali Shahrabi Farahani**[1]
[1]*School of Industrial Engineering, Iran University of Science and Technology, Tehran, Iran*
*noghondarian@iust.ac.ir, e_mohammadi@iust.ac.ir, alifarahani92@gmail.com*

**Abstract**

Proper models for prediction of time series data can be an advantage in making important decisions. In this study, we try to compare one of the most useful classic models of economic evaluation, Auto Regressive Integrated Moving Average model with one of the most useful artificial intelligence models, Adaptive Neuro-Fuzzy Inference System (ANFIS). Furthermore, we analyze the performance of these methods to predict the global gold price. Our sample data is 200 gold prices from February 2015 to October 2015. We use both methods for determination of model parameters' and to apply them on our test data. With respect to reliable evaluation methods, as root mean square of errors, it can be seen that in our test data, prediction of Adaptive Neuro-Fuzzy Inference system model is more accurate than auto-regressive integrated moving average. So we can conclude that at least in some cases where time series have non-linear trend, it is better to use Adaptive Neuro-Fuzzy Inference system for prediction.

**Keywords:** Adaptive Neuro-Fuzzy Inference System, Auto Regressive Integrated Moving Average, comparison of prediction methods, global gold price

## 1- Introduction

Prediction will provide a powerful tool for managers to be more successful in the long and short term planning for their organization. Prediction can be done in two ways: it can either be the result of deduction and analysis of an expert in a given field of knowledge, or the analysis and evaluation of raw data and statistics. In this study we consider prediction using time series data. Time series show different trends in different cases. If we want to divide this behavior into two general categories, we can say that data have either linear or non-linear trend. The purpose of this study is to analyze Auto-Regressive Integrated Moving Average model and artificial neural network model in fuzzy systems. Then with comparing these models we can conclude whether the classical Auto-Regressive Integrated Moving Average Model has the same prediction power as the neuro-fuzzy model or not. Classical model of moving average or Box-Jenkins model, have conventionally been used with data having linear trend. But in the real world there are fewer cases where data having linear trend or static state in average and variance, so recently more accurate methods of modeling non-linear systems are invented.

*Corresponding author

Two important examples of these methods are neural network and fuzzy system. Although neural network and fuzzy system have high ability for modeling of non-linear data but they also have some drawbacks. For example, to use neural network we need too much data. Sometimes lack of rules and certain tests for choosing an appropriate neural network structure causes the neural network being unable to understand the complexity of a time series observations. Hence with respect to different researches, it has been shown that we can't say with certainty that neural network has a better performance than a classical forecasting methods. In many cases especially when data show linear behavior, the classical methods predict more accurate. To fix the weaknesses of neural network models and fuzzy systems and obtain a more accurate model with higher prediction ability, we propose combining these two methods. Modeling neural networks in fuzzy system is one way of combining these two methods so that the perdition error decreases as the model continuously adapts with time series data. In this study such modeling procedure which is called Adaptive Neural-Fuzzy Inference System and Auto regressive integrated Moving Average Model are investigated and their performance are compared in a case study involving the prediction of gold price. Anticipating the gold market changes' is vital for investors so that they can be one step ahead of their opponents and have computational advantageous. Different researchers have published papers in this field. Khan (2013), used Box-Jenkins method for predicting gold price. Newer methods such as neural networks are also been used for prediction of gold price. In the work of Zahra-Nezhad and Hamid (1388), inflation rate is predicted using dynamic neural network. Some researchers tried to present a more accurate model with combination of neural networks and fuzzy systems. For example, Reuter and Moller (2010) are two German researchers who used neural networks in predicting fuzzy time series. In these papers, researchers have introduced models with high reliable results. Here we only use one model for prediction and it has not been compared with better models and weaknesses of model have not been analyzed.

## 2- Auto regressive integrated Moving Average Model

Box-Jenkins model is for identification, estimation, evaluation and prediction of single variable time series (Box et al., 2015). Single variable time series are a kind of series in which the amount of one variable in time series is related to its past amount and the amount of its present and past element of error. Processes like autoregressive, moving average and autoregressive moving average are some examples of this series.

### 2-1- Auto correlation function

$$E(X_t - \mu)(X_{t+h} - \mu) = \lambda_h \quad (1)$$

Where: $\gamma_{-h} = \lambda_{+h}$

$$\mathrm{var}(X_t) = \lambda_0 \quad (2)$$

So if h=0 we have:

### 2-2-Auto correlation function

$$\rho_h = \frac{\lambda_n}{\lambda_0} \quad ; \; \rho_h = \rho_{-h} \quad (3)$$

$$-1 \le \rho_h \le 1$$

If we divide $\lambda_h$ per $\lambda_0$ then we have ACF function like this:

So if h=0 then $P_0$=1 (Abbasi-nezhad, 1384).

## 2-3- Partial autocorrelation function

ACF (k) function in degree of k shows impure relation between $X_t$ and $X_{t-k}$. impure relation between those name is Partial Autocorrelation and is defined as simple Autocorrelation between $X_t$ and $X_{t-k}$ minus that section with linear relation between $X_t$ and $X_{t-k}$ from past have not been explained, i.e.:

$$\begin{aligned}\rho_h^* &= corr[X_t - \beta_1 X_{t-1} - \beta_2 X_{t-2} \\ &-\ldots - \beta_k X_{t-k},\ X_{t-k}]\end{aligned} \tag{4}$$

## 2-4- Difference Stationary Process (DSP)

If $X_t$ is a series like the following:

$$X_t - X_{t-1} = \beta + \varepsilon_t \tag{5}$$

It is called DSP. $\varepsilon_t$ is a stationary series with mean zero and variance $\sigma^2$ (Abbasi-nezhad, 1384).

## 2-5- Autoregressive Moving Average Process models

If we recombined two DSP with different property like AR and MA, it forms a DSP like Autoregressive Moving Average Process. For example ARMA (1,2) is written as below:

$$X_t = \alpha X_t + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} \tag{6}$$

Autoregressive Moving Average DSP is one of the most flexible patterns for single variable time series. An ARMA (p, q) series is defined as follows:

$$\begin{aligned}X_t &= \alpha X_{t-1} + \ldots + \alpha_p X_{t-p} \\ &+\varepsilon_t + \theta_1 \varepsilon_{t-1} + \ldots + \theta_q \varepsilon_{t-q}\end{aligned} \tag{7}$$

It depends on $\theta_1 \ldots \theta_q$ parameters which must be estimated. We can turn DSP (5) into an Autoregressive DSP (Abbasi-nezhad, 1384).

## 2-6- Autoregressive integrated Moving Average

If we difference a time series d times, the series is said to be integrated of order d. With autoregressive degree of P and moving average degree of qit can be shown as ARIMA (p,d, q) (Abbasi-nezhad, 1384).

## 2-7- Time series prediction of Autoregressive integrated Moving Average models

In this section we will discuss prediction by using an Autoregressive integrated Moving Average model with and without Thrust. If $\theta_0 = 0$ then we have:

$$X_t = W_t - \sum_{i=1}^{d} \binom{d}{i} (-1)^i X_{i-1} \qquad (8)$$

If $t = n+1$ and $P_n$ multiplied in each side we have:

$$P_n X_{n+l} = P_n W_{n+l} - \sum_{i=1}^{d} \binom{d}{i} (-1)^i P_n X_{n+l-1} \qquad (9)$$

We know that $\{W_i\}$ is an ARMA (p, q) with average zero and using previous method we calculate $P_n W_{n+1}$ and then $P_n X_{n+1}$ is obtained successively from above relation (khazaii, 1387).

## 3- Neuro-fuzzy system (ANFIS)

Modeling the systems with common mathematical tools like differential equations for complicated systems with uncertainty is not efficient. On the other hand fuzzy system with utilization of a set of fuzzy rules can give an accurate model for qualitative aspect of human knowledge and logical processes without using quantitative analysis. Modeling and fuzzy identification have been investigated by Takagi& Sugeno and they have obtained many practical applications in control, identification and prediction (Jang, 1993). Neuro-fuzzy network is obtained by combining fuzzy structure with artificial neural network. They can be used for identification of systems, time series prediction and other cases.

ANFIS structure that was presented in 1993 is the result of combining adaptive neural network and fuzzy inference in which hybrid training process is applied. Parameters of this ANFIS structure can be regressed for modeling systems based on available input-output data. Structures that have been presented before 1993 have less ability of adaptation in comparison with ANFIS. Furthermore After 1993 various neuro-fuzzy structures were presented. Evolving Fuzzy Neural Networks and Inference System Dynamic are two most important models of these structures. These structures (except ANFIS) somehow use clustering of data for modeling. For instance in training process of Evolving Fuzzy Neural Networks and Inference System Dynamic, new fuzzy rules are produced and clustering is done evolutionary. For this reason these structures are called “evolving”. In these networks the number and the limit of clusters change during training process (Kasabov, 2001) and (Kasabov et al., 2002).

### 3-1- ANFIS structure

ANFIS (Adaptive Network-based Fuzzy Inference System) is an adaptable and trainable network. In term of performance it is absolutely like inference fuzzy system. For simplicity we assume that our fuzzy system has two x and y inputs and an output z. Now if rules be as follows:

$Rule\,1: if\ x\ is\ A_1\ and\ y\ is\ B_1\ then\ f_1 = p_1 x + q_1 y + r_1$

$Rule\,2: if\ x\ is\ A_2\ and\ y\ is\ B_2\ then\ f_2 = p_2 x + q_2 y + r_2$ And if for defuzzification we use defuzzification of center’s average, then output will be as follows:

$$f = \frac{w_1 f_1 + w_2 f_2}{w_1 + w_2} = \bar{w}_1 f_1 + \bar{w}_2 f_2 \qquad (10)$$

$$st \quad \bar{w}_1 = \frac{w_1}{w_1 + w_2}\ , \bar{w}_2 = \frac{w_2}{w_1 + w_2}$$

The final structure equivalent to ANFIS is shown in figure 1.

Layer 1: in this layer inputs pass through membership functions:

$$O_{1,i} = \mu A_i(x) \qquad for\ i = 1,2$$
$$O_{1,i} = \mu B_i(x) \qquad for\ i = 3,4$$

Membership functions of each function could be a proper parameter that in most cases Gaussian functions are chosen. For example general form of Bell curve function is as follows:

$$\mu A(x) = \frac{1}{1 + \left| \frac{x - c_i}{a_i} \right|^{2b_i}} \tag{11}$$

Where{ $a_i, b_i, c_i$ } are set of parameters. Parameters of this layer are well known as premise parameters.
Layer2: output of this layer is the product of input signals which is actually equivalent to if part of rules.

$$Q_{2,i} = w_i = \mu A_i(x) \mu B_i(y),\ i = 1,2 \tag{12}$$

Layer 3: output of this layer is the normalization of previous layer:

$$Q_{3,i} = \bar{w}_i = \frac{w_i}{w_1 + w_2} \quad ,i = 1,2 \tag{13}$$

Layer 4:

$$Q_{4,i} = \bar{w}_i f_i = \bar{w}_i (p_i x + q_i y + r_i) \tag{14}$$

Layer 5: output of this Layer is total output of the system:

$$Q_{5,i} = \sum_i \bar{w}_i f_i = \frac{\sum_i w_i f_i}{\sum_i w_i} \tag{15}$$

Now a network is produced which is equivalent to Sugeno inference fuzzy system.

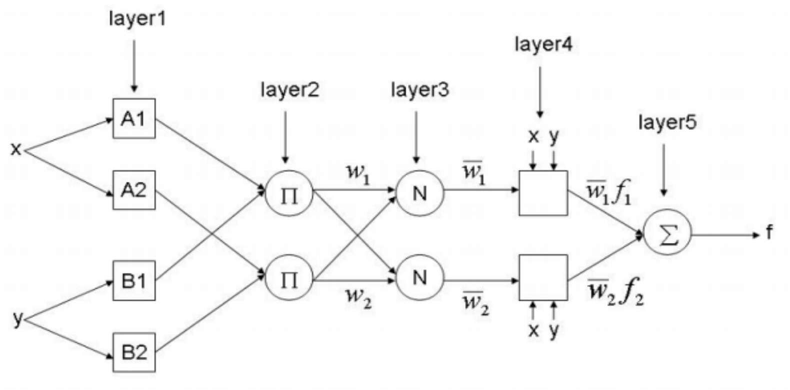


**Figure1.** Layers output

## 3-2- Clustering

One of the important issues in designing fuzzy systems is to choose proper numbers of principles. Choosing a lot of numbers for principles make the system complicated and choosing a few principles may produce weak fuzzy system which doesn't meet our purpose. In this section, number of principles is considered as an important parameter in fuzzy systems and it is determined based on input-output pair and error of the model. The main reason for clustering is grouping of input-output pairs in different categories and using a principle or a fuzzy rule for each category. The concept of clustering is to partition data into subsets and separate clusters where data in each cluster are similar as much as possible and they are very different from data in other clusters.

# 4- Case study

After explaining the theoretical aspects of these two models, we can now explain the method of modeling the global gold price data and comparing the results obtained from these models. Analysis of data with ARIMA is done by Eviews software and analysis of data with ANFIS model is done by Matlab software.

## 4-1- ARIMA model

## 4-1-1- Identification

Time series data for gold price from 2/2/2015 to 10/13/2015 is obtained from (www.kitco.com) and its plot is shown in figure 2.We use the augmented Dicky-Fuller test to check whether our data is stationary, and to conduct a unit root test , (Table 1 and **Figure3**3) (Enders, 2010) and (Khan, 2013).

As seen, the P-value is more than 0.05 and we can't reject the null hypothesis of there is a unit root, therefore our data isn't stationary. Of course this conclusion can also be reached from the autocorrelation plot which decreases very slowly. To make data stationary, we difference it once and its augmented Dicky-Fuller test is shown in figure 4. It can be seen that the null hypothesis of having unit root is now rejected with respect to a low P-value close to zero. So after first differencing, our data has become stationary. Now we can identify model using autocorrelation function and partial autocorrelation function. As can be seen, the autocorrelation function at different delays is close to zero except at delay 22 where at both plots, bars are out of bounds. Now values of p and q should be determined. These two delays should be analyzed more carefully because their values specify our model. With respect to this, there is three candidates for best model: ARIMA(0,1,22), ARIMA(22,1,0) and ARIMA(22,1,22). Now to find which model is the best, we should use Akaike and Schwartz criteria. Model with minimum Akaike and Schwartz value is considered as the best model (Akaike, 1969).

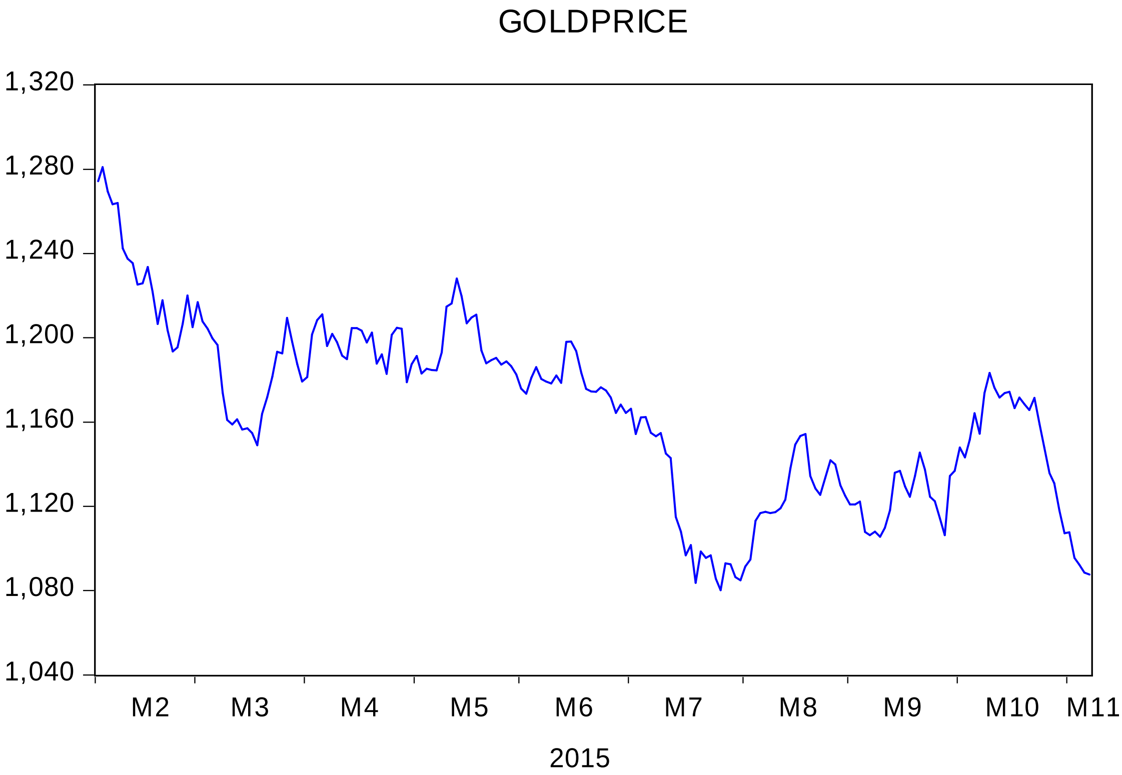


**Figure 2.** Gold price from 2/2/2015 to 10/13/2015 (www.kitco.com)

**Table 1.**Augmented Dickey-Fuller test statistic. maxlag=13

Null Hypothesis: D(GOLDPRICE) has a unit root
Exogenous: Constant
Lag Length: 0 (Automatic - based on SIC, max lag=13)

| | | t-Statistic | Prob.* |
|---|---|---|---|
| Augmented Dickey-Fuller test statistic | | -12.79072 | 0.0000 |
| Test critical values: | 1% level | -3.466786 | |
| | 5% level | -2.877453 | |
| | 10% level | -2.575332 | |

**Table 2.**Augmented Dickey-Fuller test statistic. maxlag=14

Null Hypothesis: GOLDPRICE has a unit root
Exogenous: Constant
Lag Length: 0 (Automatic - based on SIC, maxlag=14)

| | | t-Statistic | Prob.* |
|---|---|---|---|
| Augmented Dickey-Fuller test statistic | | -2.102875 | 0.2439 |
| Test critical values: | 1% level | -3.463235 | |
| | 5% level | -2.875898 | |

Date: 01/11/16 Time: 16:17
Sample: 2/02/2015 11/06/2015
Included observations: 200

| Autocorrelation | Partial Correlation | | AC | PAC | Q-Stat | Prob |
|---|---|---|---|---|---|---|
| | | 1 | 0.951 | 0.951 | 183.67 | 0.000 |
| | | 2 | 0.897 | -0.07... | 347.92 | 0.000 |
| | | 3 | 0.845 | -0.01... | 494.21 | 0.000 |
| | | 4 | 0.796 | 0.009 | 624.73 | 0.000 |
| | | 5 | 0.747 | -0.03... | 740.28 | 0.000 |
| | | 6 | 0.707 | 0.066 | 844.26 | 0.000 |
| | | 7 | 0.667 | -0.02... | 937.52 | 0.000 |
| | | 8 | 0.627 | -0.03... | 1020.2 | 0.000 |
| | | 9 | 0.590 | 0.028 | 1093.9 | 0.000 |
| | | 1... | 0.557 | 0.005 | 1159.8 | 0.000 |
| | | 1... | 0.526 | 0.007 | 1219.0 | 0.000 |
| | | 1... | 0.505 | 0.083 | 1273.7 | 0.000 |

**Figure3.** Auto correlation and partial coloration

Date: 01/11/16 Time: 16:20
Sample: 2/02/2015 11/06/2015
Included observations: 199

| Autocorrelation | Partial Correlation | | AC | PAC | Q-Stat | Prob |
|---|---|---|---|---|---|---|
| | | 1 | 0.059 | 0.059 | 0.7092 | 0.400 |
| | | 2 | 0.055 | 0.052 | 1.3293 | 0.514 |
| | | 3 | -0.03... | -0.04... | 1.5921 | 0.661 |
| | | 4 | 0.044 | 0.046 | 1.9834 | 0.739 |
| | | 5 | -0.05... | -0.05... | 2.5886 | 0.763 |
| | | 6 | 0.052 | 0.053 | 3.1430 | 0.791 |
| | | 7 | 0.099 | 0.104 | 5.2013 | 0.635 |
| | | 8 | -0.00... | -0.03... | 5.2087 | 0.735 |
| | | 9 | -0.02... | -0.02... | 5.3103 | 0.806 |
| | | 1... | -0.05... | -0.04... | 5.8898 | 0.824 |
| | | 1... | -0.07... | -0.07... | 7.0156 | 0.798 |
| | | 1... | 0.024 | 0.050 | 7.1417 | 0.848 |
| | | 1... | -0.01... | -0.02... | 7.1716 | 0.893 |
| | | 1... | -0.02... | -0.03... | 7.2965 | 0.923 |
| | | 1... | -0.04... | -0.02... | 7.6669 | 0.936 |
| | | 1... | 0.006 | 0.008 | 7.6741 | 0.958 |
| | | 1... | -0.03... | -0.01... | 7.9783 | 0.967 |
| | | 1... | -0.03... | -0.02... | 8.2787 | 0.974 |
| | | 1... | 0.015 | 0.011 | 8.3266 | 0.983 |
| | | 2... | 0.018 | 0.016 | 8.3985 | 0.989 |
| | | 2... | 0.009 | 0.010 | 8.4164 | 0.993 |
| | | 2... | -0.17... | -0.17... | 15.097 | 0.858 |
| | | 2... | 0.006 | 0.025 | 15.104 | 0.891 |
| | | 2... | -0.06... | -0.04... | 15.993 | 0.888 |
| | | 2... | 0.075 | 0.071 | 17.275 | 0.872 |

**Figure 4.** Autocorrelation and partial coloration

### 4-1-2- Evaluation

Now this question may arise that which model has better regression on data and gives better modeling. As we have mentioned before for this purpose, Akaike and Schwartz criterions (Akaike, 1969) should be used and the model that shows smaller value of these criteria gives a better regression. Considering the Akaike and Schwartz values (Akaike, 1969) for the three models, we can see these values for ARIMA (0, 1, 22) are respectively 7.262 and 7.311 which are the lowest among the three models. Therefore the parameters of our model are:

ma(22)= -0.226551
$a_0$ =-0.84820

Therefore we can define our model as:

$X_{t=}$ -0.848201+-0.226551 $X_{t-22}$

Because the q value of 22 is too high in this model indicating that 22 observations are lost in the estimation process, it is better to use p with unit value instead of q with high value. So we can substitute ARIMA (1, 1, 0) model as the best model. With respect to purpose of this study, it is possible that we have several models as best model and we can't choose definitely one model as best model. It should be noted that every model that is introduced should have scientific justification. Now for more evaluation we analyze the residuals and make sure they are not auto correlated. A residuals test in Eviews software is shown in figure 4.

**Table 3.**Convergence achieved after 4 iterations

Coefficient covariance computed using outer product of gradients

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|---|---|---|---|---|
| C | -0.848201 | 0.516620 | -1.641826 | 0.1022 |
| MA(22) | -0.226551 | 0.073713 | -3.073421 | 0.0024 |
| SIGMASQ | 80.49932 | 7.633983 | 10.54486 | 0.0000 |

| | | | |
|---|---|---|---|
| S.E. of regression | 9.040545 | Akaike info criterion | 7.262101 |
| Sum squared residue | 16019.36 | Schwarz criterion | 7.311749 |
| Log likelihood | -719.5791 | Hannan-Quinn criter. | 7.282195 |
| F-statistic | 4.685014 | Durbin-Watson stat | 1.864680 |

| Autocorrelation | Partial Correlation | | AC | PAC | Q-Stat | Prob |
|---|---|---|---|---|---|---|
| | | 1 | 0.066 | 0.066 | 0.8737 | |
| | | 2 | 0.051 | 0.047 | 1.4023 | 0.236 |
| | | 3 | -0.02... | -0.02... | 1.5067 | 0.471 |
| | | 4 | 0.027 | 0.028 | 1.6547 | 0.647 |
| | | 5 | -0.03... | -0.04... | 1.9700 | 0.741 |
| | | 6 | 0.042 | 0.045 | 2.3437 | 0.800 |
| | | 7 | 0.091 | 0.091 | 4.0581 | 0.669 |
| | | 8 | -0.03... | -0.05... | 4.3030 | 0.744 |
| | | 9 | -0.02... | -0.02... | 4.4463 | 0.815 |
| | | 1... | -0.04... | -0.04... | 4.9561 | 0.838 |
| | | 1... | -0.10... | -0.09... | 7.0873 | 0.717 |
| | | 1... | 0.019 | 0.046 | 7.1657 | 0.786 |
| | | 1... | -0.01... | -0.02... | 7.2009 | 0.844 |
| | | 1... | -0.04... | -0.06... | 7.6736 | 0.864 |
| | | 1... | -0.00... | 0.024 | 7.6736 | 0.906 |
| | | 1... | 0.012 | 0.010 | 7.7057 | 0.935 |
| | | 1... | -0.03... | -0.02... | 7.9384 | 0.951 |
| | | 1... | -0.04... | -0.02... | 8.3711 | 0.958 |
| | | 1... | 0.005 | -0.00... | 8.3760 | 0.972 |
| | | 2... | 0.021 | 0.028 | 8.4792 | 0.981 |
| | | 2... | 0.023 | 0.023 | 8.5974 | 0.987 |
| | | 2... | 0.013 | -0.00... | 8.6342 | 0.992 |
| | | 2... | -0.00... | -0.00... | 8.6353 | 0.995 |
| | | 2... | -0.04... | -0.05... | 9.1657 | 0.995 |
| | | 2... | 0.056 | 0.063 | 9.8968 | 0.995 |

**Figure 5.** Q-Statistic probabilities adjusted for ARMA term

With respect to Q-Stat values we can realize whether an error component is auto correlated. If probability value in each delay is lower than 0.05, there is autocorrelation at that delay. As can be seen, probabilities are higher than 0.05in all delays indicating there aren't any autocorrelation between error components. Also assumption of error components having normal distribution should also be investigated. Results of this test are shown in figure 4.
The probability value of J-B statistic is 0.39 which is above the 0.05significance level indicating the normality assumption can't be rejected.

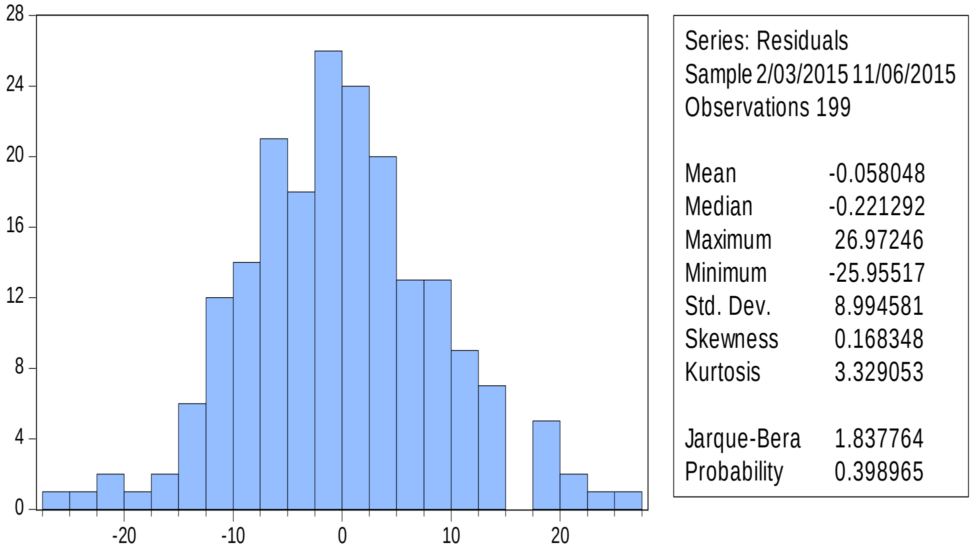


**Figure 6.** Distribution parameters

**4-1-3- Results of ARIMA model**

The results of 40 days forecast using ARIMA (1, 0, 22) model is shown in figure 7. To evaluate forecasts we should consider Mean Absolute Error (MAE), root mean square error (RSME) and mean absolute percentage error (MAPE). The lower value of these statistics determine better forecast. In the next step we will also do these forecasts with ANFIS model and then we will compare the results.

## 4-2- Modeling and forecasting with ANFIS model

As we have mentioned before in this method errors decrease continuously until it is adapted with system. With respect to type of data and to reach lower level of error it is better to use clustering which here in this example clustering algorithm is based on error data. Membership function in fuzzy system is also Gaussian.

Now we separate data into two Train and Test groups. At first, based on Train data we find proper model and then in order to see whether our model has good ability to forecast the future, we compare future forecast with Test data. So we use first 160 data for model identification and model parameter estimation. Plots are shown in **Error! Reference source not found.**.

With respect to first 160 data which is Train data, model parameters are estimated as: $\mu$=0.00040101, $\sigma$=0.084453. To see whether the model has good future forecasting ability we test it with our next 40 data.

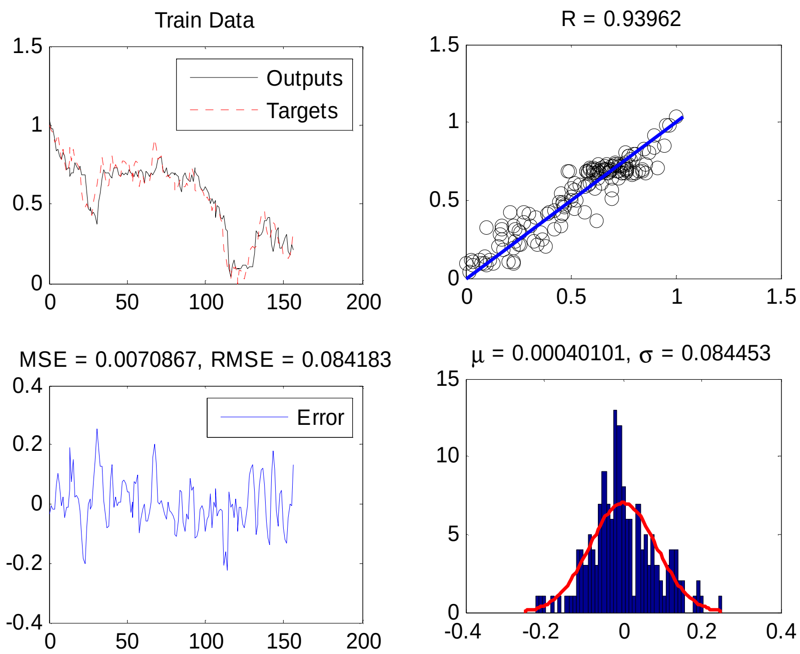


**Figure 7.** First behavioral analysis

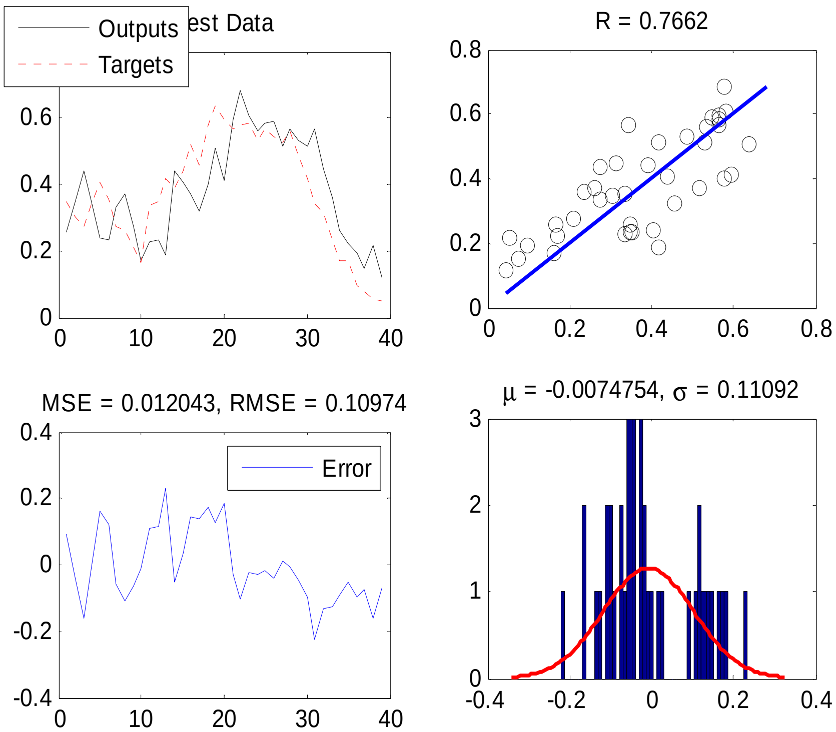


**Figure 8.** Second behavioral analysis

#### 4-2-1- Results from ANFIS model forecast

Results are shown in **Error! Reference source not found.**. Other error values are as follows:

MAPE =0.3061

MAE=0.092149

Comparing real data with forecasted data shows that our model has high forecasting ability with RSME, MSE, MAE and MAPE show very small values. This testifies that our model can forecast future with small error.

### 4-3- Summary and comparison

To compare the two models, we should compare error criterions like RSME, MSE, MAE and MAPE as shown in table 3. The Model that shows a lower value is better. It is clear that all forecast errors in ANFIS model is much lower than those errors in ARIMA model. So we can say with high certainty that ANFIS model gives better forecasts than ARIMA model.

## 5- Conclusion

With respect to results obtained in this study, we can conclude that ANFIS model can forecast future gold price much better than ARIMA model. But it is not right to conclude that with every kind of data, the difference between errors of the two methods will be as high as that is reported in this model. Indeed type of data will affect the modeling of time series. In this study, our ARIMA data are close to non-linearity rather than to be linear and so gave us proper model but the data are unable to give a model with small amount of error. Maybe if data type is closer to linearity, ARIMA model can give more accurate forecast. However, it is mentioned in introduction. So depending on type of data, test result will be different but main point in this study was high modeling ability of ANFIS method. If we look at the structure of this method carefully, we understand that modeling procedure in this method is so powerful and efficient because in this method error value is calculated continuously until approaching to minimum possible, we can reduce error to any desired amount. So this method has identification ability of complicate and dynamic data. Especially when we partition data with clustering, it helps us to have similar data in each cluster and it can easily identify relation between data and reduce error more and more. We already knew that neural network and fuzzy system have high ability in forecasting of time series when they are non-linear. Combination of these two methods can generate a new synergy and has more advantages than applying them separately for forecasting the time series.

**Table 4.** ARIMA model versus ANFIS model

| | RMSE | MAE | MAPE |
|---|---|---|---|
| ARIMA model | 49.87085 | 42.34575 | 3.665439 |
| ANFIS model | 0.10974 | 0.092149 | 0.3061 |

## References

Abbasi-nezhad, H.,(1384). Advanced econometric. University of Tehran.

Akaike, H. (1969). Fitting autoregressive models for prediction. *Annals of the institute of Statistical Mathematics*, *21*(1), 243-247.

Box, G. E., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time series analysis: forecasting and control*. John Wiley & Sons.

Enders, Walter (2010). *Applied Econometric Time Series (Third ed.). New York: Wiley. pp. 206–215. ISBN 978-0470-50539-7.*

Jang, J. S. (1993). ANFIS: adaptive-network-based fuzzy inference system. *IEEE transactions on systems, man, and cybernetics*, *23*(3), 665-685.

Kasabov, N. K., & Song, Q. (2002). DENFIS: dynamic evolving neural-fuzzy inference system and its application for time-series prediction. *IEEE transactions on Fuzzy Systems*, *10*(2), 144-154.

Kasabov, N. (2001). Evolving fuzzy neural networks for supervised/unsupervised online knowledge-based learning. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, *31*(6), 902-918.

Khan, M. M. A. (2013). Forecasting of gold prices (Box Jenkins approach). *International Journal of Emerging Technology and Advanced Engineering*, *3*(3), 662-670.

khazaii, M (1387). Introduction to time series analysis by S plus software, Tehran, Iran statistical center, statistical institute

Reuter, U., & Möller, B. (2010). Artificial neural networks for forecasting of fuzzy time series. *Computer‐Aided Civil and Infrastructure Engineering*, *25*(5), 363-374.

Wang, L. X. (1997). A Course in Fuzzy Systems and Control, Prentice-Hall PTR. *Englewood Cliffs, NJ*..

Zahra-nezhad, M.,Hamid,Sh. (1388).Inflation rate Prediction of iran economy by using artificial time dynamic neural networks (time series point of view), quantitative economic journal ,6:1.