

Forecasting Gold Price using Data Mining Techniques by Considering New Factors

A. Hatamlou^{1*} and M. Deljavan²

1. Department of Computer Science, Khoy Branch, Islamic Azad University, Khoy, Iran.
2. Department of Computer Science, Urmia Branch, Islamic Azad University, Urmia, Iran.

Received 21 August 2017; Revised 03 February 2018; Accepted 26 April 2018

*Corresponding author: hatamlou@iaukhoy.ac.ir (A. Hatamlou).

Abstract

Gold price forecast is of great importance. Many models have been presented by researchers to forecast gold price. It seems that although different models could forecast gold price under different conditions, the new factors affecting gold price forecast have a significant importance and effect on the increase in the forecast accuracy. In this work, different factors were studied in comparison with the previous studies on gold price forecast. In terms of time span, the collected data was divided into the three groups of daily, monthly, and annually. The conducted tests using new factors indicated an accuracy improvement of up to 2% in the neural networks methods, 7/3% in the time series method, and 5/6% in the linear regression method.

Keywords: Gold price Forecast, Data Mining, Time Series, Neural Networks, Regression.

1. Introduction

Gold has been considered as a precious metal since the past. Thus the prediction of its price has been of great importance.

New advancements and innovations like the modern manufacturing technologies, expansion of operations, and thousands of other important factors have made the economy of nations complicated and expanded. Investment forms an important part of the country. Today, one of the most important and favorite subjects of economists and financial analysts is to explain the process of price fluctuations that have created different methods and attitudes in this regard. Meanwhile, the forecast of these fluctuations is not possible easily according to the lack of accurate data about the factors affecting the fluctuations of the gold market. Using quantitative methods in financial markets and investment has

become more necessary than before. In international contracts and exchanges, the ability of gold price forecast can reduce the risk due to the fluctuations of gold price. However, the forecast of gold price fluctuations is still difficult.

In the recent years, the use of artificial intelligence technologies to solve these problems has been of great interest. Neural networks are one of the methods used in financial forecasts, and have usually provided a better answer than the previous methods. Gold price forecast is essential due to its economic dimension. On the other hand, it has been always considered as an appropriate solution to serve and keep the capital among people who have always attempted to recognize the factors affecting gold price and increase their profit.

2. Review of literature

Mirmirani and Li have used back-propagation neural networks (BP) by genetic algorithms in order to

prevent the fluctuations of gold price, and have proved that it depends on the fluctuations of gold price for a short time [1]. Using genetic algorithms to find the architecture of optimal networks minimizes the reliance on modeling the network selection. They have used the only intended characteristic for gold price forecast, i.e. gold price, from 12/31/1974 to 12/31/1998.

In [2], the authors have used Adaptive Neuro-Fuzzy Inference System (Anfis) and compared it with other time series forecast methods like (ARMA) for a proper process of forecast. In fact, they studied the proposed models to forecast the process of stock price process for the next day for gold. The results have indicated that Anfis has good results in the percent of accuracy using the strategy of buy and hold and other methods. In this work, only the characteristic of daily gold price was used.

Leyla sarfaraz and amir afsar have used the fuzzy-neural networks model to forecast Iran's gold price and compared its forecast to the regression model. They concluded that the method of fuzzy-neural networks model has a high accuracy in forecast. In this work, the characteristics of Iran's gold price, global gold price, inflation, and TEPIX index were used for forecast [3].

In [4], the authors have developed the two forecast models of multi-linear regression and ARMA model to forecast Thailand's gold price. They used the characteristics of Thailand's gold price and global gold price to prove that ARMA (1,1,1) was one of the most appropriate methods. They also proved that the simultaneous use of multi-linear regression and ARMA model increased the accuracy of forecast in a short time. The idea that ARMA can forecast gold price has been accepted in many countries. For example, in Australia, Selranathan compared the forecast of London's daily gold price at the Center for Economic Research by the ARMA model and proved that ARMA cost much less and was effective enough in gold price forecast.

The characteristics used in this forecast were the Australian dollar index, American dollar index, oil price, Japanese Yen, Pound index, and bank interest rate.

Dunis and Nathani have predicted the daily gold and silver price using advanced regression analyses and different linear and non-linear models [5]. Their major goal was to find a daily quantitative business strategy. Authors used the ARMA models like linear model for comparative objectives by non-linear

models like the nearest neighbors, multi-layer perceptron, and higher order neural networks (MLP). The obtained results indicated that non-linear models like MLP and HONN were much better models. In fact, the authors confirmed that non-linear models could be used effectively in creation of an extra return in those markets. They used the gold and silver price indices for prediction from May 2000 to January 2007.

In [6], the authors have argued that the multiple linear regression (MLR) model is better for the forecast of future gold price. MLR model is based on the economic factors that affect gold price. The first model considered all possible independent variables, and the second one considered only some independent variables (commodity Research Bureau Index, Euro/commodity exchange rate, and Treasury bill) during the forecast. Finally, the second model reached a high level of forecast accuracy than the first one.

Shafiee and Topal in [7] have used the modified economic models in long-term processes, reverting jump and dip diffusion models for gold price forecast. They used two types of forecast statistical errors like MAE and RMSE to compare the accuracy of the models. They also used ARIMA models for gold price forecast. Neural networks were equipped with genetic algorithms, which had the advantages of non-linear model simulation when there was a small previous knowledge on the structure of problems. The studies indicate that these systems provide a better forecast when they are compared with the traditional economic models and used the characteristics of global gold price, inflation rate, and oil price.

In [8], the authors have used the ARIMA method (James Box) for gold price forecast, and studied a variety of time series models. They concluded that the time series method with the values (0,1 and 1) was the best model with a minimum of errors, and used the characteristics related to the data of gold price from January 2003 to March 2012.

In [9] the authors have used the two methods of Garch and neural networks for global price forecast and concluded that neural networks method was more efficient than the Garch method, and also used the characteristics of Turkey's gold price from February 2010 to January 2014.

Moradi et al, have used the method of gmdh neural networks for Iran's gold price forecast, and compared it with mlp neural networks [10]. The good

and acceptable results obtained from the criteria of evaluating GMDH neural network performance and the high ability of this network are recognizing the patterns dominating the data and also the unique characteristics such as rapid convergence, high accuracy, and its ability. In this work, seven characteristics affecting the forecast were studied, which were silver price, stock price index, inflation rate, oil price, global gold price, dollar index, and bank interest rate.

3. Factors and techniques used

The determination of accurate and efficient characteristics is essential in data mining, and the degree of forecast accuracy depends on appropriate data and the algorithm used in it. The characteristics studied in this article are the consumer price index that is usually as a driver for investment in the gold market as an anti-shock against inflation. Gold price has a tendency towards displacement with cpi, and since the USA currency and Eurozone currency are the two main currencies of the global gold markets, both the United States consumer price index and Eurozone consumer price index were used. Also, Iran's consumer price index was used in the forecast of Iran's gold price. Another factor involved is SPDR that is the largest gold investor in the world and thus the purchase of SPDR gold stock directly means the increase in demand and gold price. Open interest refers to unfinished interests or contracts, and like the previous factor, the more amounts of open interest means more demand for gold and increase in gold price. USD is the balancing factor of gold price or the index that negatively affects the increase in gold price. This index determines the performance of the USD against foreign currencies. Dollar fluctuations mainly affect gold price because currently, the global gold market is dominated by the USD. The debilitation of dollar increases gold price because investors want a lower risk of exchange rate.

The official sales center where a number of the largest gold owners officially announce that buying and selling gold at any time negatively affects the demand of the gold market and finally, reduce the gold market price.

Another effective factor is GDP in the USA and Iran that consists of some factors like personal consumption expenditure, government consumption expenditure, gross capital and import and export rate which will be discussed below. Economic growth leads to a higher oil consumption, and finally, the

increase of oil price. This effect is shown with the positive relationship between GDP and increases the prices of oil and gold. The only factor involved to reduce oil price is the control of oil reserves in the USA. Personal consumption expenditure refers to the final household consumption and the expenditure of good/service consumption by resident families whether inside or outside the economic realm. In case of the goods and services whose preparation cost is regarded as the household final consumption expenditure, the increase in personal consumption expenditure reduces the growth of economy and increases gold price.

The government consumption expenditure and gross capital include the payment of the expenditure for good/service prices in the realm of government activities or include the good/service prices spent on the realm of government activities. The increase in government consumption expenditure reduces the growth of economy and increases gold price.

Good/service export leads to the economic development of countries, and has a significant effect on GDP and increases prices. Good/service import, led to economic growth spurt in the developed and developing countries. Some studies show that the effect of import on economic growth is uncertain in some countries which probably lead to the structure of export in these countries. This, import is among the growth tools in the modern world. The increase in labor expenditure is another factor that increases oil price that leads to the increase in gold price.

Another effective factor is the oil reserves of the USA. As the oil reserves become less than its previous level, the market will notice the tension in supply, and oil price will increase resulting in the increase in gold price. Thus the factor of oil reserves has a positive relationship with the output flow of oil price reduction.

Housing price is another factor involved in gold price. Any type of facility provided for currency and coin affects the price of housing because all of them are in need of each other. Gold price increases with the increase in housing price. However, gold price is not necessarily reduced with the decrease in gold price. Also the demand for gold purchase and its price increases with the decrease in the number of exchanges in housing.

Some countries like China and India, purchased gold due to the reduction of dollar value in the global market. Also the increase in demand increased the price of this precious metal in the global market.

Thus the increase in gold demand in India and China increased gold price in the world.

The inflation due to the use of expansionary monetary policies of central banks was another factor affecting the increase of precious metals and gold price with the increase in liquidity volume and its injection to the financial system. Thus the expansionary monetary policies and interest rate reduction had a lot of effect on the increase in cheap loan demands on behalf of low-income people, and made the financial system of America use the creation power of its credit money more than ever, so that it led to the increase in price levels and uncertainty in the global economy in the process of economic evolutions, the creation of bank credit, and the use of expansionary monetary policies by monetary and financial policy-markets in the USA. The increase in the U.S. government's debt and the concerns that were created about the reduction of dollar value in the world alongside the European debt crisis (especially the euro-zone countries), the reports on negative credit rating in Greece and Spain, the increase in concerns on the depth of debts in the Eurozone increased the continuity of precious metals and gold demand, turned the precious metals market into a safe place for the assets of investors, and also increased the global price of precious metals. The increase of international concerns than the U.S. government's debt, as a country that has the biggest economy in the world and has also a considerable GDP, reduces the value of dollar and increases the global demand for the purchase of gold.

Another effective factor in gold price is interest rate. People do not have a tendency towards the investment of their capitals with the reduction of interest rate in western countries, etc. Thus they tend to purchase gold with the decrease in tendency towards keeping dollar since gold has the feature of value stability even under the conditions of international wars, crises, and tensions.

The growth of liquidity or in other words, the volume of money, total money and quasi money affect the creation of inflation and increase in gold price.

Since the amount of goods and services is limited in the society, the amount of liquidity should be, to an extent, equal to the goods and services. If the amount of liquidity increases, the goods and services in the society will reduce, and prices will increase. Liquidity control is one of the most important demands of countries. Thus different policies are developed for this purpose, for example, the increase

in technology and inland production that prevents inflation by collecting sufficient commodities. The relationship between monetary base, money, and liquidity is as follows:

Monetary base x Multiplier = liquidity= money
 Monetary base (High powered money) equals the net assets or net debts of the central bank that are given on the balance sheet of this bank and includes:

Foreign assets, bill, coin, public sector debt, public companies, and banks' debts.

Another factor affecting gold price is the injection of the revenue obtained from selling oil to the society increasing liquidity, inflation, and gold price. Excessive government borrowing from the central bank is another factor that imposes inflation to the society. In order to control the current inflation rate, the effect of government borrowing from the banking system should be proportioned to the supply and demand curve of funds and financial resources.

The most important reason to increase prices in Iran is the state budget deficit. When the previous state provides the deficit through creating liquidity and increases liquidity before the end of year, it causes inflation and increases gold price.

One of the indices of Tehran stock exchange that was studied in the current study is the total price index: Tehran stock exchange calculated and released its price index called TEPIX since March 1990. This index included 52 companies listed in the stock exchange at that time. This index is calculated using Equation 1.

$$TEPIX_t = \frac{\sum_{i=1}^n P_{it} Q_{it}}{D_t} * 100 \tag{1}$$

P_{it} : Stock price of the i -th company at time t
 D_{it} : Base number at time t that equals $\sum P_{i0}$ at Q_{i0} at the source time

p_{i0} : Stock price of the i -th company at the source time

Q_{i0} : Number of the stocks of the i -th company at time t

P_{it} : Number of the stocks released by the i -th company at the source time

N : Number of the companies including the index

Tehran stock exchange price index includes all companies listed in the stock, and if a corporate sign is closed or not exchanged for a while, the price of its last exchange will be considered in the index. There is a significant relationship between macroeconomic variables, Tepix, GDP, explanation of consumer price index, gold price, Rial to Euro

currency exchange rate, global oil price, and liquidity.

One of the factors affecting gold price is copper price because there is 75% gold in 18 carat gold that is combined with a base metal like nickel or copper. There is also 58/5% gold in 14 carat gold. The higher the carat of gold is, the softer the metal is. 14 carat gold has more resistance and durability than the other kinds of gold because it is combined with another metal that strengthens it. As copper price is higher, it increases gold price, and as copper is lower, it decreases gold price. In previous studies, the only factors that were studied in Iran’s gold price forecast were the USA 24 carat gold, oil price, dollar price, Tepix, Iran’s 18 carat gold, and consumer index.

In general, the data was collected from different sources since 2009 until the end of the year 2013, and classified in 55 fields and 3 tables as daily, monthly, and annually. They were considered as the input of data mining algorithms whose output is the field of gold price forecast. In this study, Clementine software version 12 made by SPSS company and also different methods of data mining like regression, neural networks, and time series were used [11-17]. In multi-layer perceptron (MLP) neural networks, there are two issues: the selection of an appropriate architecture and the selection of an appropriate training algorithm. Appropriate algorithm means the optimal selection of the number of layers, the number of neurons at each layer, and the type of activation function of each neuron. The optimal architecture of neural networks is based on datasets and their features. Different training algorithms are used in neural networks. The most common training algorithm of these networks is back-propagation algorithm. In this algorithm, the new calculated output value is composed of the real value at any step, and the network’s weights are modified according to the obtained error. So that, the amount of the obtained error at the end of each iteration is less than the error obtained at the previous iteration. To achieve the forecast error, we use MAE (mean absolute error), RMSE (root-mean-square error), and R² (the square of the correlation coefficient). Their formulas are as follow:

$$MAE = \frac{1}{n} \sum_{i=1}^n |x_{obs,i} - x_{model,i}| \tag{2}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (x_{obs,i} - x_{model,i})^2}{n}} \tag{3}$$

$$r = \frac{\sum_{i=1}^n (x_{obs,i} - \bar{x}_{obs,i}) \cdot (x_{model,i} - \bar{x}_{model,i})}{\sqrt{\sum_{i=1}^n (x_{obs,i} - \bar{x}_{obs,i})^2 \cdot \sum_{i=1}^n (x_{model,i} - \bar{x}_{model,i})^2}} \tag{4}$$

In the above equations, X_{obs,i} is the i-th ratio of the observed gold price, X_{model,i} is the i-th ratio of the predicted gold price, $\bar{X}_{obs,i}$ is the total mean of the observed gold price ratio, and $\bar{X}_{model,i}$ is total mean of the predicted gold price ratio.

Forecast by using regression analysis or regression analysis is a statistical technique to study and model the relationships between the variables. Regression analysis is one of the mostly used methods among the statistical techniques. Correlation coefficient was used to determine the type of degree of the relationship between a quantitative variable and another one. Correlation coefficient is another criterion used in determining the correlation between the two variables. This coefficient is between 1 and -1, and also equals zero in case of no relationship between the two variables. Correlation between the two random variables X and Y is defined as follows:

$$corr(X, Y) = \frac{cov(X, Y)}{\sigma_x \sigma_y} = \frac{E[(X - \mu_x)(Y - \mu_y)]}{\sigma_x \sigma_y} \tag{5}$$

in which E is the mathematical expected operator, cov is covariance, corr is Pearson correlation, and σ is standard deviation.

In forecast method, first, the stationary of the series is measured using the ARIMA method. To ensure the stationary of the series, the Dickey - Fuller Extended Test is conducted by the software. If there is no stationary in the series, the software performs the test by differencing for once. In this step, the value of T test should be less than the critical value, as shown in the table. In the next step, the model and the type of its order were identified. To do this, AFC and PACF figures are used that, respectively, show autocorrelation function and differentiated partial autocorrelation function. Then the estimate and tests of the determined models are divided into two parts: accepted models and non-accepted models. The models have two necessary conditions for being accepted. First, their coefficients should be significant, and second, their critical value should be less than 0/05. Finally, the models are studied using Topsis algorithm, and the best model is selected.

4. Experimental results

Table 1. The factors used in the previous studies.

Consumer index in Iran	carat gold price in the USA	carat gold price in Iran	Dollar price	Oil price	Tepix
------------------------	-----------------------------	--------------------------	--------------	-----------	-------

Table 2. The factors used as daily in the model.

The value of exchanges in SPDR	The amount of gold in SPDR	carat gold price in the USA	carat gold price in Iran	Oil price
Tepix	euro / dollar Exchange rate	Copper price	Dollar price	The number of transactions in SPDR

The data that was collected monthly is as follows:

Table 3. The factors used as monthly in the model.

Inflation in USA	Inflation in Iran	Consumer price index in USA	Consumer price index in Iran
	Labor expenditure index in USA	Labor expenditure index in Iran	Consumer price index in Europe

The data that was collected annually is as follows:

Table 4. The factors used as annually in the model.

The public sector debt to the central bank	The volume of imports in Iran	personal consumption expenditure in the USA
three-year interest rate	The state budget deficit	two-year interest rate
The volume of exports in the USA	personal consumption expenditure in Iran	Four-year interest rate
The volume of imports in the USA	China's demand rate	Monetary base
The central bank's net foreign assets	India's demand rate	Money multiplier
The volume of exports in Iran	Index of gross fixed capital formation in the USA private sector	Changes in the monetary base
Five-year interest rate	The us. Government's debt	The trade balance in the USA
One-year interest rate	Index of gross fixed capital formation in iran private sector	The amount of bank debt to the central bank
Oil reserves in the USA	Gross fixed capital formation in the private sector	Foreign exchange revenues from oil sales
The estate transactions	government consumption expenditure and iran's gross capital	government consumption expenditure and America's gross capital
GDP in Europe	GDP in Iran	GDP in the USA

In this table, the test and train data was separated from each other and subtracted from the data of the test set as 5%. Then, they were added to the train data, and the neural networks of MLP model were implemented using Clementine software. The

accuracy of forecasts can be seen in the following table. In the MLP method, the number of input neurons equals 55, the number of hidden layer neurons equals 9, and one neuron in the outer layer.

Table 5. The forecast accuracy using the MLP method.

Train data	Test data	Accuracy of forecast using factors in this research	Accuracy of forecast using factors in previous studies
%۵۰	%۵۰	99.4	94.1
%۵۵	%۴۵	99.5	96.5
%۶۰	%۴۰	99.4	96.6
%۶۵	%۳۵	99.5	96.6
%۷۰	%۳۰	99.4	96.6
%۷۵	%۲۵	99.4	97.5
%۸۰	%۲۰	99.4	97.3

In the following methods, the error rate in neural networks is calculated using Equations 2, 3, and 4.

Table 6. The obtained error values in MLP model using the factors in previous studies.

R SQUARE	MAE	RMSE
97.5	3253.014	5032.12

Table 7. The obtained error values in MLP model using factors in this work.

R SQUARE	MAE	RMSE
99.5	1984.392	2023.142

Forecast using regression method

In table 8, the correlation coefficient value and the relationships between variables and gold price were obtained using Equation 5. As the coefficient is

more, it shows the strong relationship between the intended variable and gold price. Lower coefficients show a weak relationship between the intended variable and gold price.

Table 8. The correlation coefficients related to the factors used in the model.

Euro / Dollar Exchange rate 0.34	The volume of imports in Iran 0.47	Inflation in Iran 0.82
The amount of gold in SPDR 0.4	The state budget deficit 0.84	Copper price 0.3
The volume of exports in the USA 0.75	personal consumption expenditure in Iran 0.64	Oil price 0.5
The number of transactions in SPDR 0.5	The volume of exports in Iran 0.57	Inflation in the USA 0.82
The value of exchanges in SPDR 0.4	Consumer price index in Iran 0.9	Dollar price 0.5
The volume of imports in the USA 0.37	Consumer price index in USA 0.81	Tepix 0.71
The public sector debt to the central bank 0.82	India's demand rate 0.31	Money multiplier 0.75
personal consumption expenditure in the USA 0.5	China's demand rate 0.68	Changes in the monetary base 0.42
Oil reserves in the USA 0.64	Labor expenditure index in USA 0.56	Foreign exchange revenues from oil sales 0.45
GDP in Europe 0.37	GDP in Iran 0.84	Labor expenditure index in Iran 0.3
The estate transactions 0.59	Index of gross fixed capital formation in the USA private sector 0.5	¥carat gold price in the USA 0.91

In the following tables, the error value in the regression method of Enter model was calculated using Equations 2, 3, and 4.

Table 9. The obtained error values in regression method using the factors in previous studies.

R SQUIRE	MAE	RMSE
92.8	6631.243	7023.62

Table 10. The obtained error values in regression method using the factors in this work.

R SQUIRE	MAE	RMSE
98.5	3264.339	4025.65

The obtained regression equation is equal to:
 The price of Iran’s 18 carat gold= Oil price * 34.2 + the USA 24-carat gold price * 39.21 +The number of transactions in SPDR * 0.071+ consumer index in the USA* * (-228.9) + inflation rate in the USA * (-1504.6) + Consumer index in Iran* 905.9+ bank interest rate* 50.25+ inflation rate in Iran* (-527.8) + demand arte in India *(65.35) + the amount of the U.S government’s debt to the federal government*

(-0.027) + China's foreign exchange reserves* (0.00617) + dollar price* * (-1.108) + copper price* (2903.7) + GDP in Iran* (47.66) + EUR / USD exchange rate * (15924.2) + Tepix* (-0.66) + volume of imports and exports* *8.03) + GDP in the USA* (30.63) + China’s demand rate* (90.36) + revenue from oil sales*(112).

Forecast using ARIMA time series

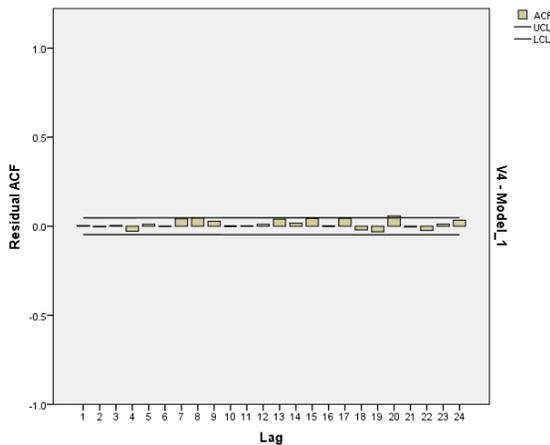


Figure 1. The differentiated auto-correlation function.

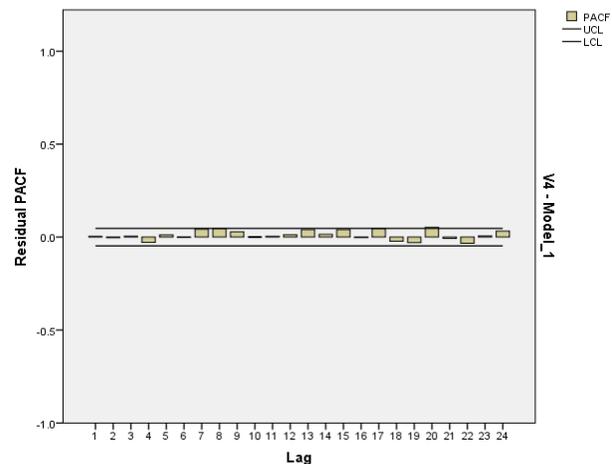


Figure 2. The differentiated partial auto-correlation function.

In this step, first, the stationary of the series is measured. To ensure the stationary of the series, the Dickey - Fuller Extended Test is conducted by the software. If there is no stationary in the series, the software performs the test by first-order differencing. In this step, the value of T test should be less than the critical value, as shown in the table. In the next step,

the model and the type of its order were identified. To do this, AFC and PACF figures are used that, respectively, show autocorrelation function and differentiated partial autocorrelation function. The figures show autocorrelation function and differentiated partial autocorrelation function respectively.

According to the figures, these two functions for the autoregressive order are equal to $p=1, 2, 3, 4$ and $q=1, 2, 3, 4, 5$. Thus 28 models are selected for the model. In the next step, the determined models are estimated and tested, and the models are divided into two parts: accepted models and non-accepted models. The models have two necessary conditions for being

accepted. First, their coefficients should be significant, and second, their critical value should be less than 0/05. According to the model software (2,1, and 6) the most appropriate model is selected. In the following tables, the error value in time series method was calculated by Equations 2, 3, and 4.

Table 11. The obtained error value in ARIMA model using the factors of this work.

Model	Number of Predictors	Model Fit statistics		
		square R	RMSE	MAE
V4	5	0.981	5009.513	3514.357

Table 12. The obtained error value using the time series method of ARIMA model by studying the factors in previous studies.

Model	Number of Predictors	Model Fit statistics		
		square R	RMSE	MAE
V4	1	0.908	10945.394	8312

The obtained parameters of ARIMA model are as follow:

As it can be observed, the value 1 in tables 13 and 14 means that the series does not have stationary and the software used first-order differentiating to make the

series stationary and also the critical value of threshold t in both tables is acceptable and significant according to the software that estimated it.

Table 13. The parameters of ARIMA by studying the studied factors in work.

V4	Natural Logarithm			Estimate	T
		AR	Log 2	-0.81	-3.335
		Difference		1	
		MA	Log16	-0.71	-2.927

Table 14. the parameters of ARIMA by studying the factors in previous studies.

V4	Natural Logarithm			Estimate	T
		AR	Log 2	-0.72	-3.003
		Difference		1	
		MA	Log16	-0.73	-3.039

In tables 15 and 16, the results related to data mining techniques were examined by studying the factors in

this work and previous studies. The results can be studied and compared with each other.

Table 15. The comparison between techniques using factors in this work.

R SQUIRE	MAE	RMSE	Method name	Model name
99.5	1984.392	2023.142	MLP	Neural network
98.5	3264.339	4025.65	Enter	Regression
98.1	3514.307	5009.513	ARIMA	Time series

Table 16. The comparison between techniques using the factors in previous studies.

R SQUIRE	MAE	RMSE	Method name	Model name
97.5	3253.014	4032.12	MLP	Neural network
92.8	6631.243	7023.62	Enter	regression
90.8	8312	10945.394	ARIMA	Time series

5. Conclusion

The most important factors affecting Iran’s gold price are global gold price, oil price, consumer index in Iran, the US Government’s debt, inflation in Iran, GDP in Iran, bank interest rate, and Tepix, among which inflation is the most important factor. Chronic inflation in the country and inflation expectations due to the increase in the volume of money in circulation increases gold price. In this study, the characteristics that were more effective than previous studies were studied. It was concluded that the accuracy of forecast, in case of using the studied factors in this work is higher than the factors that were already considered. Also, the accuracy of forecast in neural networks is higher than regression and time series methods. Regression method is better than the time series method. Also we had the maximum improvement in time series method with 7/3% increase in accuracy.

References

[1] Mirmirani, S. & Li, H.C. (2004). Gold price, neural networks and genetic algorithm, Computational Economics, vol. 23, no. 4, pp. 193–200.

[2] Atsalakis, G. S., Skiadas, C. H. & Braimis, I., (2007). Probability of Trend Prediction of Exchange Rate by ANFIS, in Recent Advances in Stochastic Modeling and Data Analysis, World Scientific, vol. 5, no. 3, pp. 414-422.

[3] Sarfaraz, L. & Afsar, A. (2005). A study on the factors affecting gold price and a neuro-fuzzy model of forecast, Tarbiat Modares Economic Research Journal, vol. 16, no. 2, pp. 70-75.

[4] privat Khaemasunun, P. (2007). Forecasting Thai Gold Prices, College of Innovation, Thammasat University, Thailand, vol. 23, no. 3, pp. 96–103.

[5] Dunis, C. L. & Nathani, A. (2007). Quantitative trading of gold and silver using nonlinear models, Neural Network World, vol. 16, no. 2, pp.93–111.

[6] Ismail, Z., Yahya, A. & Shabri, A. (2009). Forecasting gold prices using multiple liner regression method, American Journal of Applied Sciences, vol. 6, no. 8, pp. 1509–1514.

[7] Shafiee, S. & Topal, E. (2010). An overview of global gold market and gold price forecasting, Resources Policy, vol. 35, no. 3, pp.178–189.

[8] Massarrat Ali Khan, M. (2013). Forecasting of Gold Prices (Box Jenkins Approach), International Journal of Emerging Technology and Advanced vol. 3, no. 3, pp. 662-670.

[9] Habip, K. & Turgut, U. (2014). Forecasting the Gold Returns with Artificial Neural Network and Time Series, International Business Research, vol. 7, no. 11, pp. 139.

[10] Moradi, H., Jokar, I. & Forouzantaba, A. (2015). Modeling and forecasting gold price using gmdh neural network, Indian Journal of Fundamental and Applied Life Sciences, vol. 5, no. 1, pp. 30-41.

[11] Kul Kavani R. V. & Devale A. B., (2012). A Review of Data Mining Techniques in, vol.1, no.2, pp. 150-156.

[12] Mingming, T. & Jinliang, Z. (2012). A multiple adaptive wavelet recurrent neural network model to analyze gold prices, Journal of Economics and Business, vol. 64, no. 2, pp.275–286.

[13] Lin, K & Yeh, C. (2012). Use of Data Mining Techniques, International Journal of Engineering and Technology Innovation, vol. 2, no. 2, pp. 42-53.

[14] sadatrasoul, S., gholamian, M., Siami, M. & Hajimohammadi, Z. (2013). Credit scoring in banks and financial institutions via data mining techniques: A literature review, Journal of AI and Data Mining, vol. 1, no. 2, pp. 119-129.

[15] Moslehi, F., Haeri, A. & Moini, A. (2018). Analyzing and Investigating the Use of Electronic Payment Tools in Iran using Data Mining Techniques, Journal of AI and Data Mining, vol. 6, no. 2, pp. 417-437.

[16] Sedighi, Z. & Boostani, R. (2018). Extracting Prior Knowledge from Data Distribution to Migrate from Blind to Semi-Supervised Clustering, Journal of AI and Data Mining, vol. 6, no. 2, pp. 287-295.

[17] Izakian, Z. & Mesgari, M. (2015). Fuzzy clustering of time series data: A particle swarm optimization approach, Journal of AI and Data Mining, vol. 3, no. 1, pp. 39-46.

پیش‌بینی قیمت طلا با استفاده از تکنیک‌های داده‌کاوی با در نظر گرفتن عوامل جدید

عبدالرضا حاتم‌لو^{۱*} و مینو دلجووان^۲

^۱ گروه کامپیوتر، واحد خوی، دانشگاه آزاد اسلامی، خوی، ایران.

^۲ گروه کامپیوتر، واحد ارومیه، دانشگاه آزاد اسلامی، ارومیه، ایران.

ارسال ۲۰۱۷/۰۸/۲۱؛ بازنگری ۲۰۱۸/۰۲/۰۳؛ پذیرش ۲۰۱۸/۰۴/۲۶

چکیده:

پیش‌بینی قیمت طلا از اهمیت زیادی برخوردار است. مدل‌های زیادی توسط محققان برای پیش‌بینی قیمت طلا ارائه شده است. به نظر می‌رسد که اگر چه مدل‌های مختلف می‌توانند قیمت طلا را در شرایط مختلف پیش‌بینی کنند، عوامل جدیدی که پیش‌بینی قیمت طلا را تحت تاثیر قرار می‌دهند اهمیت زیادی دارند و بر افزایش دقت پیش‌بینی تاثیر می‌گذارند. در این کار، عوامل مختلف در مورد پیش‌بینی قیمت طلا در مقایسه با مطالعات قبلی مورد مطالعه قرار گرفته است. از لحاظ بازه زمانی، داده‌های جمع‌آوری شده به سه گروه روزانه، ماهانه و سالانه تقسیم شده‌اند. آزمایش‌های انجام شده با استفاده از عوامل جدید نشان‌دهنده بهبود دقت به میزان ۲٪ در روش شبکه‌های عصبی، ۷٫۳٪ در روش سری زمانی و ۵٫۶٪ در روش رگرسیون خطی می‌باشد.

کلمات کلیدی: پیش‌بینی قیمت طلا، داده‌کاوی، سری زمانی، شبکه‌های عصبی، رگرسیون.