JSEG

JOURNAL OF SMART ECONOMIC GROWTH

www.jseg.ro ISSN: 2537-141X Volume 5, Number 1, Year 2020

Fuzzy Auto-Regressive Integrated Moving Average (FARIMA) Model for Forecasting the Gold Prices

Sahed Abdelkader, Mekidiche Mohammeed, Kahoui Hacen

Department of Economics Sciences, the University center of Maghnia, Algeria

sahed14@yahoo.fr , mkidiche@yahoo.fr , hacenkahwi@gmail.com

Abstract: *In this study, the Fuzzy Auto-Regressive Integrated Moving Average (FARIMA) Model has been used to predict gold prices, The main objective was to estimate the fractional parameters by using the fuzzy regression method of TANAKA. The prediction accuracy of the FARIMA method was measured and compared with the ARIMA method using mean square error (MSE) and mean square error (RMSE) for the time period from 2010 to 2018. Gold prices were also predicted using the two methods for the year 2019. The research concluded that FARIMA models are more efficient than ARIMA models in predicting gold prices.*

JEL classification: G1, G17, C5, C22

Key words: Gold Prices, Fuzzy Set, ARIMA, FARIMA

1. Introduction

The Gold has gained dominant powers over the economy as a tool for exchanging goods and services, as the actual circulation of gold dates back to the fourteenth century, with gold in the 19th and 20th centuries serving as a cash guarantee for the issue of banknotes. Later, gold trading joined other addresses such as commodities, stocks, and others.

The Gold is considered the most invested metal, as investors buy gold to distribute commercial risks. Gold affects the global economy, either as a factor in the country's prosperity or a factor in the deterioration of the country's economy, so gold has a major impact on the American economy, which is the largest buyer of gold, the more the dollar has deteriorated whenever the price of gold increases, which leads investors to buy gold in the hope of Gold will preserve their assets.

The prediction of fluctuations in gold prices It is a big issue given its economic and political effectiveness. With this in mind, predicting more precisely the volatility of gold prices is important for commodity markets and the global economy. The importance of forecasting gold prices to help makers of monetary policy and investment decisions better.

The Many studies also confirmed a strong relationship between gold prices and oil prices, where they connected that the high price of gold has a powerful relationship with oil prices.

From this standpoint, many models are designed to predict the gold price, for example ARIMA models, neural network models, genetic algorithm, and fuzzy logic.

This study seeks to fill that gap, and in this context this study was divided into four sections, the section touched upon a simple introduction in which we show the importance of the importance of the topic and the second section was subjected to the most important previous studies in this topic and the third section in which the methodology of the study was presented It is represented in the ARIMA models and the fuzzy ARIMA models. The fourth section includes the application of the two models in predicting gold prices. The fifth section deals with a study comparing the used models with reliance on both MAE and RME.

2. Literature review

In recent years, authors have increased their interest in modeling, forecasting and analyzing time series of economic and financial variables. There are many studies that focus on predicting gold prices, including the following:

In 2015, the authors Kristjanpoller, W., et al, published an article entitled, Gold price volatility: A forecasting approach using the Artificial Neural Network–GARCH model. They expanded the scope of expert systems, prediction, and model by applying an artificial neural network (ANN) to the GARCH method to generate ANN-GARCH. They applied the hybrid ANN-GARCH model to predict gold volatility . The results showed a general improvement in prediction using ANN - GARCH compared to the GARCH method alone.

In 2012, The authors Yazdani, A., et al. published an article entitled, Forecasting gold price changes by using adaptive network fuzzy inference system. They used adaptive neuro-fuzzy inference system (ANFIS) and artificial neural network (ANN) model to estimate the price of gold, and compare it with The traditional statistic is represented by the ARIMA (autoregressive integrated moving average) model. Three performance measures, the coefficient of determination (R2), root mean squared error (RMSE), mean absolute error (MAE), were used for performance values different from the developed methods. The results showed that the ANFIS model outperformed the other methods (ARIMA, ANN).

Khan, M. M. A. (2013). Developed a Forecasting of gold prices (Box Jenkins approach). They used the gold price data from January 2, 2003 to March 1, 2012 as the study period, the data was divided into two parts, one until January 2 to build the model while the rest was used to predict the price of gold and verify the accuracy of the model, they used the Box-Jenkins method, to build the model. they obtained that the ARIMA (0,1,1) model is the most appropriate to use in predicting the price of gold, relying on this at the expense of Root Mean Square Error.

Guha, B., & Bandyopadhyay, G. (2016). published an article entitled, Gold price forecasting using ARIMA model. They applied ARIMA model to the time series of the gold price on historical data from November 2003 to January 2014, an analysis of the price performance of gold 10 years ago gave them the value traded in MCX, ARIMA model (1, 1, 1) that helps them predict future values for gold. ARIMA (1, 1, 1) was chosen from six different model parameters because it provided the best model that met all the criteria of fit stats while five others failed to match stats.

Yaziz, S. R., et al. (2013). published an article entitled, The performance of hybrid ARIMA-GARCH modeling in forecasting gold price. They examined the performance of hybrid of the ARIMA models with the GARCH models in analyzing and forecasting daily gold price data series. The Box-Cox formula is used in the data transformation step to address non stationarity in variance. Results of 40-day gold price data series that the hybrid ARIMA (1,1,1) -GARCH (0,2) model yielded better and more accurate compared to the ten previous methods of forecasting in literatures.

3. Methodology and data

3.1. ARIMA Models

The ARIMA model has been used as a control group to compare the performance of forecasts. The linear ARIMA models are the most common and general, with the ARIMA models having an AR component called the autoregressive, and an MA component called the moving average. Where ARIMA adopts the following equation (1):

$$Y_t = \varphi_{t-1}Y_{t-1} + \varphi_{t-2}Y_{t-2} + \cdots\cdots + \varphi_{t-p}Y_{t-p} + C + \varepsilon_t - \theta_{t-1}\varepsilon_{t-1} - \theta_{t-2}\varepsilon_{t-2} - \cdots\cdots \theta_{t-q}\varepsilon_{t-q} \qquad (1)$$

The parameters of forms p and q are determined by the autocorrelation function (ACF), the partial autocorrelation function (PACF) and the Akaike criterion information (AIC) (Zhang, Y., et al., 2019).

When the p and q values have been determined by comparing several models, following is the estimation step where the maximum likelihood method is used.

Following the estimation of several models , the diagnostic step consists of validating the model in the prediction process, to this end, the series residuals are examined and subjected to several tests to ensure the integrity of the model with respect to problems arising from error autocorrelation and the Hetroscedasticity problem.

Once we diagnostic the model, there are two cases. First, if the model is Best, we go immediately to the last step that is forecasting. If the model is not good, then we do the feedback process by modifying the B and S parameters, and then we make sure the model is good until we reach the best model, See figure (1) :

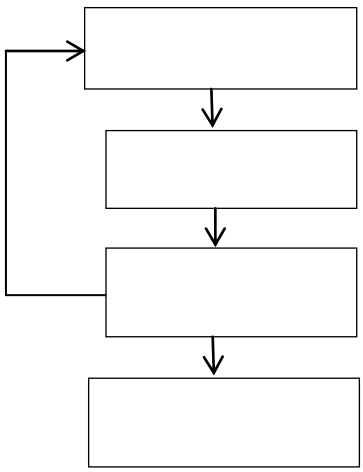


Fig. 1 – ARIMA Processes

3.2. The Fuzzy Autoregressive Integrated Moving Average (FARIMA) model

The ARIMA model has parameters $\varphi_1, \varphi_2, \cdots, \varphi_p$ and $\theta_1, \theta_2, \cdots \theta_q$ are crisp, But in reality these parameters are not crisp but are fuzzy parameters which are denoted by $\tilde{\varphi}_1, \tilde{\varphi}_2, \cdots \tilde{\varphi}_p$

and $\tilde{\theta}_1, \tilde{\theta}_2, \cdots \tilde{\theta}_q$. The parameters are in a triangular fuzzy numbers, which are used in Fuzzy Autoregressive Integrated Moving Average (FARIMA) model. The FARIMA (p, d, q) model is described by a fuzzy function with fuzzy parameters are defined by (Mehdi Khashei,. et al,. 2015):

$$\tilde{W}_t = \tilde{\varphi}_1 W_{t-1} + \cdots \tilde{\varphi}_p W_{t-p} + a_t - \tilde{\theta}_1 a_{t-1} - \cdots - \tilde{\theta}_q a_{t-q} \qquad (2)$$

Eq. (2) is modified as:

$$\tilde{W}_t = \tilde{\beta}_1 W_{t-1} + \cdots + \tilde{\beta}_p W_{t-p} + a_t - \tilde{\beta}_1 a_{t-1} - \tilde{\beta}_2 a_{t-2} - \cdots - \tilde{\beta}_q a_{t-q} \qquad (3)$$

The triangular fuzzy numbers were used to define the fuzzy parameters as follows:

$$\mu_{\tilde{\beta}_i}(\beta_i) = \begin{cases} 1 - \dfrac{|\beta_i - \alpha_i|}{c_i} & \text{if } \alpha_i - c_i \le \beta_i \le \alpha_i + c_i \\ 0 & \text{otherwise} \end{cases} \qquad (4)$$

Where:

$\mu_{\tilde{\beta}_i}(\beta_i)$:is the membership function of the fuzzy set that represents parameter β_i.

α_i : is the center of the fuzzy number.

c_i : is the width or spread around the center of the fuzzy number.

The membership function of the fuzzy dependent variable takes the following form:

$$\mu_{\tilde{W}}(W_t) = \begin{cases} 1 - \dfrac{\left|W_t - \sum_{i=1}^{p} \alpha_i W_{t-1} - a_t + \sum_{i=p+1}^{p+q} \alpha_i a_{t+p-i}\right|}{c_i} & \text{for } W \neq 0\,, a_t \neq 0 \\ 0 & \text{otherwise} \end{cases} \qquad (5)$$

However, to solve FARIMA parameter estimation problems, this model has been transformed into a linear programming problem as follows:

$$\text{Minimize } S=\sum_{i=1}^{p}\sum_{t=1}^{q} c_i|\varphi_{ii}||W_{t-i}|+\sum_{i=p+1}^{p+q}\sum_{t=1}^{k} c_i|\rho_{i-p}||a_{t+p-i}|$$

subject to

$$\sum_{i=1}^{p}\alpha_i W_{t-i}+a_t-\sum_{i=p+1}^{p+q}\alpha_i a_{t+p-i}+(1-h)(\sum_{i=1}^{p} c_i|W_{t-1}|+\sum_{i=p+1}^{p+q} c_i|a_{t+p-i}|\geq W_t \qquad (6)$$

$$\sum_{i=1}^{p}\alpha_i W_{t-i}+a_t-\sum_{i=p+1}^{p+q}\alpha_i a_{t+p-i}-(1-h)(\sum_{i=1}^{p} c_i|W_{t-1}|+\sum_{i=p+1}^{p+q} c_i|a_{t+p-i}|\leq W_t$$

$t=1,2,\dots\dots..k$

$c_i\geq 0$ for all $i=1,2,\dots.p+q$

Where:

S: is the total sum of the fuzziness

h : is the threshold value

p : is the degree of AR(p)

q : : is the degree of MA(q)

φ_{ii} : is the partial autocorrelation coefficient of time lag i

ρ_{1-p} : p is the autocorrelation coefficient of time lag $i-p$

The above linear programming issue is called the Tanaka model relative to the author Tanaka, who discovered this method. To solve the linear programming problem, an LNDO program was used to solve this problem (fahdel, et al. 2009).

The process for the FARIMA model is given as follows (Fang-Mei Tseng,. Et al. 2001):

First step: find the optimum solution for parameters α *=(α1* ,α2*,…….,α*p+q) of ARIMA(p,d,q) model.

Second step: Using the formulation (6) to determine the minimal fuzziness.

So the FARIMA model is:

$$\widetilde{W}_t=\langle\alpha_1,c_1\rangle W_{t-1}+\cdots+\langle\alpha_p,c_p\rangle W_{t-p}+a_t-\langle\alpha_{p+1},c_{p+1}\rangle a_{t-1}-\cdots$$
$$-\langle\alpha_{p+q},c_{p+q}\rangle a_{t-q} \qquad (7)$$

Third step: We delete the data around the model's upper bound and lower bound when the fuzzy ARIMA model has outliers with wide spread.

3.3. Performance Evaluation

In order to evaluate the performance of the methods used in this study, each of the two was used Coefficient of Determination (R2), Mean Squared Error (MSE) and Root Mean Squared Error (RMSE), It is shown in the following table (1) (Le, T. T., et al., 2020).

Table 1 - validation criteria

Code	Definition	Equation
(R2)	Coefficient of Determination	$R^2 = \frac{SSE}{SST}$
(MSE)	Mean Squared Error	$MSE = \frac{\sum(Y_t - \hat{Y}_t)^2}{N}$
(RMSE)	Root Mean Squared Error	$RMSE = \sqrt{\frac{\sum(Y_t - \hat{Y}_t)^2}{N}}$

3.4. Data source

In this study, monthly data for gold prices were used USD / grams, for the period from 01/01 to 12/12/2019. In total 120 data for 120 months. Where this data was obtained from the World Bank website. The data are divided into two parts: (i) in-sample period for the first 108 observations; (ii) out-of-sample for the last 12 observations.

3.5. Data descriptive

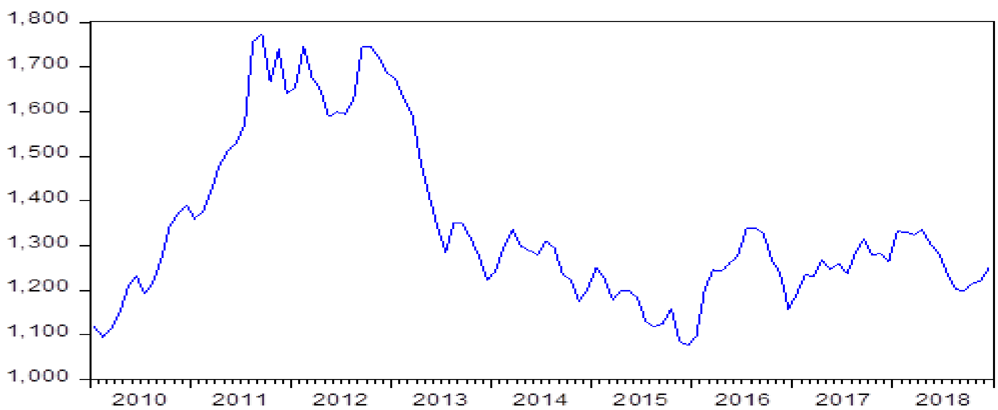


Fig. 2 – Monthly gold price

Through the previous figure of the gold price series, note that it has a random walk pattern, with a trend and a seasonal pattern.

4. Results analysis

First step: Determine the optimal values for the parameters in the ARIMA model

For this, we use the ADF test to determine whether the gold price series is stationary or not. The ADF test values are disclosed in Table (1).

Table 1: ADF Test of GP

Null Hypothesis: GP has a unit root			
Exogenous: Constant			
Lag Length: 0 (Automatic - based on SIC, maxlag=12)			
		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-1.564983	0.4969
Test critical values:	1% level	-3.492523	
	5% level	-2.888669	
	10% level	-2.581313	
*MacKinnon (1996) one-sided p-values.			

So, the probability value of t statistic is greater than 5%, so the seies is not stationary at the original level. , And since the series is of the type DS., The difference Technique has been used to convert it stationary. So the series is stationary at the first level, and therefore the seies is integrated from the first order and shown in Table (2), where we note that the probability value is less than 5%.

JOURNAL OF SMART ECONOMIC GROWTH

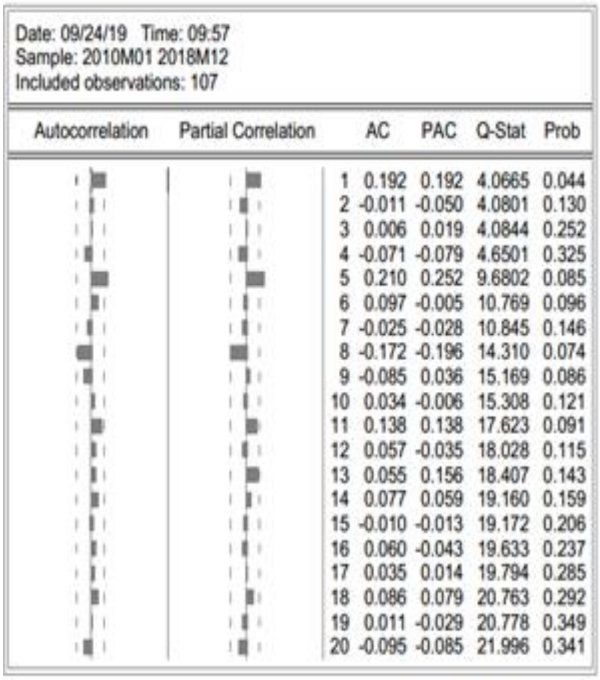
JSEG

www.jseg.ro ISSN: 2537-141X Volume 5, Number 1, Year 2020

Table 2: ADF Test of D(GP)

		t-Statistic	Prob.*
Null Hypothesis: D(GP) has a unit root			
Exogenous: Constant			
Lag Length: 0 (Automatic - based on SIC, maxlag=12)			
Augmented Dickey-Fuller test statistic		-8.383538	0.0000
Test critical values:	1% level	-3.493129	
	5% level	-2.888932	
	10% level	-2.581453	
*MacKinnon (1996) one-sided p-values.			

As for the identification phase, the correlogram of the autocorrelation and partial correlation was used for the series of differences, Figure (3) illustrates this:



Date: 09/24/19 Time: 09:57
Sample: 2010M01 2018M12
Included observations: 107

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.192	0.192	4.0665	0.044
		2	-0.011	-0.050	4.0801	0.130
		3	0.006	0.019	4.0844	0.252
		4	-0.071	-0.079	4.6501	0.325
		5	0.210	0.252	9.6802	0.085
		6	0.097	-0.005	10.769	0.096
		7	-0.025	-0.028	10.845	0.146
		8	-0.172	-0.196	14.310	0.074
		9	-0.085	0.036	15.169	0.086
		10	0.034	-0.006	15.308	0.121
		11	0.138	0.138	17.623	0.091
		12	0.057	-0.035	18.028	0.115
		13	0.055	0.156	18.407	0.143
		14	0.077	0.059	19.160	0.159
		15	-0.010	-0.013	19.172	0.206
		16	0.060	-0.043	19.633	0.237
		17	0.035	0.014	19.794	0.285
		18	0.086	0.079	20.763	0.292
		19	0.011	-0.029	20.778	0.349
		20	-0.095	-0.085	21.996	0.341

Fig. 3 – Correlogram of D(GP)

Through the previous figure, and based on the Akaike and Schwarz criteria, it was found that the model that gives the lowest value to these criteria is the ARIMA (1.1.0) model.

After we got to know the model, you are now doing the estimation process and Table (3) shows that, by examining the model, it becomes clear that the AR parameter is significantly different from zero.

Table 3: Estimate the model

Dependent Variable: D(GP)
Method: ARMA Generalized Least Squares (Gauss-Newton)
Date: 09/24/19 Time: 10:12
Sample: 2010M02 2018M12
Included observations: 107
Convergence achieved after 4 iterations
Coefficient covariance computed using outer product of gradients
d.f. adjustment for standard errors & covariance

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1.248294	5.602478	0.222811	0.8241
AR(1)	0.193360	0.095924	2.015769	0.0464

$$D(GP)_t = 1.248 + (0\,.\,193)D(GP)_{t-1} \ldots \ldots (8)$$

Second step: Using the formulation (6) to determine the minimal fuzziness.

In the first step, we set the optimum parameters (α1, α0) = (1.248, 0.193), Which is considered as input for the second step, and for the purpose of extracting fuzzy parameters, we using the formulation (6), with (h = 0) which appears in equation (9). The results were to forecast gold prices for the year 2019 shown in the following table:

$$\widetilde{W}_t = 1.248 + (0\,.\,193, 0.821)W_{t-1} \ldots \ldots (9)$$

Table 4: Forecasting goal price

Date	ARMA(1,1,0)	FARIMA LB	FARIMA UB
2019-1	1257,38975	1247,25844	1258,73561
2019-2	1259,98677	1246,87551	1262,61699
2019-3	1261,736	1247,02499	1265,6387
2019-4	1263,3216	1247,27724	1268,4945
2019-5	1264,87562	1247,54931	1271,31828
2019-6	1266,42354	1247,82521	1274,13587
2019-7	1267,97029	1248,10186	1276,95228
2019-8	1269,51682	1248,37864	1279,76845
2019-9	1271,0633	1248,65545	1282,58458
2019-10	1272,60977	1248,93227	1285,4007
2019-11	1274,15623	1249,20908	1288,21682
2019-12	1275,7027	1249,4859	1291,03294

In the third step, there is no data about the upper limit of the model and the lower limit the ARIMA fuzzy model has a lower spread.

In order to measure the accuracy of the prediction of the FARIMA method and its comparison with ARIMA models, the mean square error (MSE) and root mean square error (RMSE) for the two models were obtained. Table (5) shows that:

Table 5: Performance Evaluation

RMSE	MSE	Models
15,1469932	229,431402	FARIMA
16,0502852	257,611654	ARIMA

It is clear from the above results that the FARIMA model is the best in forecasting gold prices, and this according to the predictive accuracy measures represented by MSE and RMSE as the values of these measures are less than what is in the ARIMA model.

5. Conclusions

Through the study of the topic of forecasting gold prices, a set of conclusions were reached:

- The time series of gold prices is not stationary in the level.
- The time series of gold prices is stationary in the first order.
- By examining AFC and PAFC, it was found that the type is AR (1).

After comparing the FARIMA and ARIMA models and relying on MSE and RMSE precision measures, the FARIMA model outperformed the ARIMA model in forecasting gold prices.

CONFLICTS OF INTEREST AND PLAGIARISM: The authors declare no conflict of interest and plagiarism.

REFERENCES

1. Kristjanpoller, W., & Minutolo, M. C. (2015). *Gold price volatility: A forecasting approach using the Artificial Neural Network–GARCH model*. Expert Systems with Applications, 42(20), 7245-7251.
2. Yazdani-Chamzini, A., Yakhchali, S. H., Volungevičienė, D., & Zavadskas, E. K. (2012). *Forecasting gold price changes by using adaptive network fuzzy inference system*. Journal of Business Economics and Management, 13(5), 994-1010.
3. Khan, M. M. A. (2013). *Forecasting of gold prices (Box Jenkins approach)*. International Journal of Emerging Technology and Advanced Engineering, 3(3), 662-670.
4. Guha, B., & Bandyopadhyay, G. (2016). *Gold price forecasting using ARIMA model*. Journal of Advanced Management Science, 4(2).
5. Yaziz, S. R., Azizan, N. A., Zakaria, R., & Ahmad, M. H. (2013, December). *The performance of hybrid ARIMA-GARCH modeling in forecasting gold price*. In 20th International Congress on Modelling and Simulation, Adelaide (pp. 1-6).
6. Zhang, Y., Yang, H., Cui, H., & Chen, Q. (2019). *Comparison of the Ability of ARIMA, WNN and SVM Models for Drought Forecasting in the Sanjiang Plain*, China. Natural Resources Research, 1-18.
7. Mehdi Khashei, Mohammad Ali Montazeri, Mehdi Bijari (2015). *Comparison of Four Interval ARIMA-base Time Series Methods for Exchange Rate Forecasting*. I.J. Mathematical Sciences and Computing, , 1, 21-34.
8. Fang-Mei Tseng, Gwo-Hshiung Tzeng, Hsiao-Cheng Yu, Benjamin J.C. Yuan (2001). *Fuzzy ARIMA model for forecasting the foreign exchange market*. Fuzzy Sets and Systems 118 9-19.
9. Le, T. T., Pham, B. T., Ly, H. B., Shirzadi, A., & Le, L. M. (2020). *Development of 48-hour Precipitation Forecasting Model using Nonlinear Autoregressive Neural Network*. In CIGOS 2019, Innovation for Sustainable Infrastructure (pp. 1191-1196). Springer, Singapore.

10. fahdel, A. T., Najla, S. EL. (2009). *Fuzzy Autoregressive Model With an Application*. IRAOI JOURNAL OF STATISTICAL SCIENCES.