Journal of Mathematics Technology and Education

Vol. 1, No. 1, 2021 | 11 - 18

Journal of Mathematics Technology and Education

Gold Price Forecasting Using Autoregressive Integrated Moving Average (ARIMA) Method

Ruth Elizabeth[1], Syahriol Sitorus[2]

[1] *Student of Mathematics Department, Faculty of Mathematics and Natural Sciences, Universitas Sumatera Utara*

[2] *Mathematics Department, Faculty of Mathematics and Natural Sciences, Universitas Sumatera Utara*

Abstract. Gold is one type of investment that is starting to be favored by many people. Every investment has risk, including gold investment. This risk of loss needs to be minimized to obtain maximum profit. So it takes proper planning and forecasting to invest in gold. This study aims to determine the best forecasting model for the price of 1 gram gold bullion produced by PT Aneka Tambang (ANTAM), using the ARIMA method. The data used in this study comes from historical data in 2016 to 2019 to predict gold prices in 2020-2021. Gold Price Forecasting using the ARIMA method produces ARIMA(1,1,0) with the model model $X_t = X_{t-1} + 0.09265 - 0.1578(X_{t-1} - \emptyset X_{t-2}) + e_t$. . With error values MAE= 94491.42, MAPE= 10.47776409, MSE= 12281027973, RMSE=110819.7996.

Keyword: ARIMA, Forecasting, Gold Price, Time Series

Abstrak. *Emas merupakan salah satu jenis investasi yang mulai banyak digemari oleh masyarakat. Setiap investasi memiliki resiko, termasuk investasi emas. Resiko kerugian ini perlu diminimalisir untuk memperoleh keuntungan maksimal. Maka diperlukan perencanaan dan peramalan yang tepat untuk melakukan investasi emas. Penelitian ini bertujuan untuk menentukan model peramalan terbaik dalam harga emas batangan 1 gram produksi PT Aneka Tambang (ANTAM), dengan metode ARIMA. Data yang digunakan dalam penelitian ini berasal dari data historis di tahun 2016 sampai 2019 untuk meramalkan harga emas di tahun 2020-2021. Peramalan Harga Emas menggunakan metode ARIMA menghasilkan ARIMA(1,1,0) dengan model* $X_t = X_{t-1} + 0.09265 - 0.1578(X_{t-1} - \emptyset X_{t-2}) + e_t$. *Dengan nilai error MAE= 94491.42, MAPE= 10.47776409, MSE= 12281027973, RMSE=110819.7996.*

Kata Kunci: *ARIMA, Harga Emas, Peramalan,* Time Series

Received 10 Oct 2021 | Revised 19 Oct 2021 | Accepted 15 Nov 2021

*Corresponding author at: Mathematics Department, Faculty of Mathematics and Natural Science, University of North Sumatera, Medan 20155, Indonesia

E-mail address: rutherlizabeth@gmail.com , syahriol@usu.ac.id

e-ISSN: xxxx-xxxx | DOI: 10.32734/jomte.v1i1.6859
Journal Homepage: https://talenta.usu.ac.id/jomte

1 Introduction

Gold is one that has a low risk. Putri, et al. in [1], the annual risk of gold is 13.66%, under silver investment of 30.27%, the Indonesian Stock Market (JKSE) of 40.59% and several property investments such as the Kemis Market with a risk of 14.74% and Sukarmantri 19.51%.

The size of the risk faced in investing is shown by the size of the deviation of the expected rate of return with the actual rate of return [2]. Risk is also often associated with deviations from the accepted and expected outcomes. To calculate risk, the method that is widely used is the standard deviation which measures the absolute deviation of the value that has occurred with the expected value [3].

Investment is the investment of a number of funds made at this time in the hope of obtaining profits in the future [4]. The higher the benefits offered, the higher the risk that must be borne. Several types of investments, such as property, stocks, and gold, have become very popular, starting from the small to the upper class because they can be used as additional income.

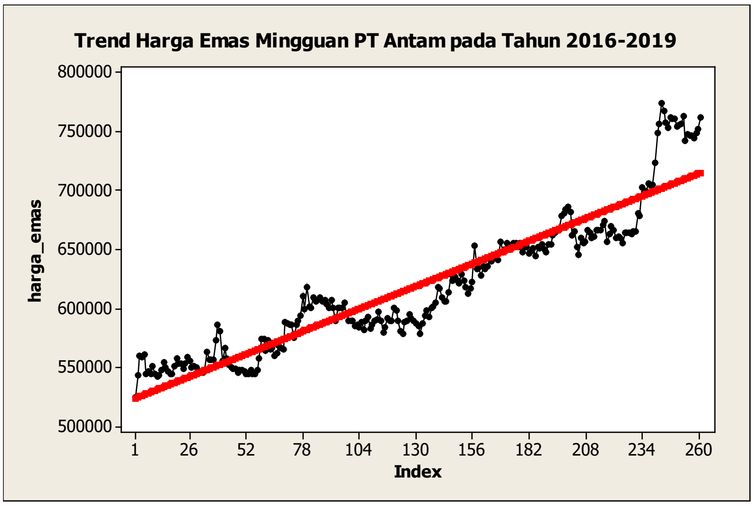

Figure 1 PT Antam's weekly price trend in 2016 - 2019

The forecasting model used in forecasting the gold price is the times series or time series, using the Autoregressive Integrated Moving Average (ARIMA) method. Method invented by George Box and Gwilym Jenkins. The ARIMA method makes full use of past and present data to produce accurate short-term forecasts. The ARIMA method will work well if the time series data used are statistically dependent or related to each other.

To minimize losses and reduce investor doubts about the weekly gold investment value, this study will provide information in the form of modeling and predictions of PT Antam's gold price in the future based on previous price data.

2 Literature Review

Stationary

One of the data stationarity tests is through the unit root test. This test was developed by David Dickey and Wayne Fuller as the Augmented Dickey-Fuller (ADF) Test.

$$ADF = \frac{\hat{\delta}}{SE(\hat{\delta})} \tag{1}$$

Differencing

Eliminating unstationary in a time series can be done by differencing, using the formula:

$$X'_t = X_t - X_{t-1} \tag{2}$$

Autocorrelation Function (ACF)

Autocorrelation Function is a function that shows the increase between observations in a time range with observations of the previous time. The equation to determine the coefficient of ACF

$$r_k = \frac{\sum_{t=1}^{n-k}(X_t - \bar{X})(X_{t+k} - \bar{X})}{\sum_{t=1}^{n}(X_t - \bar{X})^2} \tag{3}$$

Partial Autocorrelation Function (PACF)

The Partial Autocorrelation Function is a function that shows the magnitude of the partial correlation between observations at time t and observations in the previous time range. The equation for determining the PACF coefficient:

$$\phi_{kk} = \frac{r_k - \sum_{j=1}^{k-1} \phi_{k-1,j} r_{k-j}}{1 - \sum_{j=1}^{k-1} \phi_{k-1,j} r_j} \tag{4}$$

Autoregressive Integrated Moving Average (ARIMA) Model

With the general ARIMA model (p,d,q) for the simple case of ARIMA (1,1,1), as follows:

$$(1 - B)(1 - \emptyset_1 B)X_t = \mu' + (1 - \theta_1 B)e_t \tag{5}$$

Mean Absolute Error (MAE)

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|e_i| \tag{6}$$

Mean Absolute Percentage Error (MAPE)

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{\hat{X}_t - X_t}{X_t}\right) \times 100\% \tag{7}$$

Mean Absolute Error (MSE)

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(\hat{X}_t - X_t\right)^2 \tag{8}$$

Root Mean Square Error (RMSE)

$$RMSE \cong \sqrt{\frac{\sigma_Y^2[1-\rho_{X,Y}]}{N}\left\{2+\left[1+\rho_{X,Y}\right]\left[\frac{X_i-\mu_X}{\sigma_X}\right]^2\right\}} \tag{9}$$

3 Result and Discussion

Seasonal Test

Table 1 Experiment Seasonal Test

Source of Variation	Free Degrees	Number of Squares	Sum of Squares Average	Test Statistic
Average	1	636387.3301	636387.3301	
Between Season	3	85226487986974.80	28408829328991.6	4.3530
In Season	205	19578575693871.8	6526191897957.28	
Total	209	104805064317234		

From the table above, it is found that $F_{table} < F_{count}$ is $2.65 < 4{,}365$, then H_0 is accepted, where the weekly gold price does not have a seasonal element.

Stationary Data Against Mean

The price of gold every week shows an increase or decrease and is not constant against the average. To find out more precisely about the stationarity of the data, it is also necessary to test the stationarity with the ADF test.

```
        Augmented Dickey-Fuller Test

data:  harga_emas
Dickey-Fuller = -1.474, Lag order = 5, p-value = 0.7965
alternative hypothesis: stationary
```

Figure 2 Augmented Dickey-Fuller Test before differencing

Because the p-value = 0.7965 < 0.05 alpha value, the Ho hypothesis is rejected so that the data value is declared not stationary on average.

Differencing

The first difference from the transformation results to the variance can be done with the equation $X_t = X_t - X_{t-1}$ for t = 2, 3, …, 209. Then the following results are obtained:

$X'_2 = X_2 - X_1 =$ 128.2701831-128.5292151- = 0.258121475

$X'_3 = X_3 - X_2 =$ 128.2701831-128.5292151 = -0.258121475

…

$X'_{209} = X_{209} - X_{208} =$145.0836-144.381= 0.700119

```
        Augmented Dickey-Fuller Test

data:  diff_emas
Dickey-Fuller = -6.1976, Lag order = 5, p-value = 0.01
alternative hypothesis: stationary
```

Figure 3 Augmented Dickey-Fuller Test after differencing

After differencing, the ADF value in the first differentiation has a p-value which indicates the data is stationary on average.

Temporary ARIMA Identification ACF and PACF plots

The following is the ACF and PACF plot of weekly gold prices after the first differencing

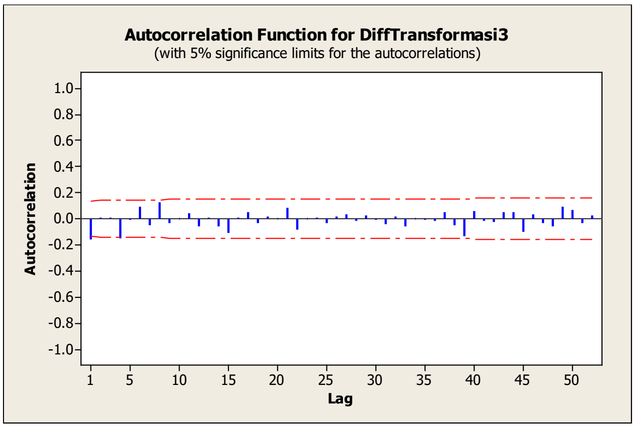

Figure 4 Autocorrelation plot after first differencing

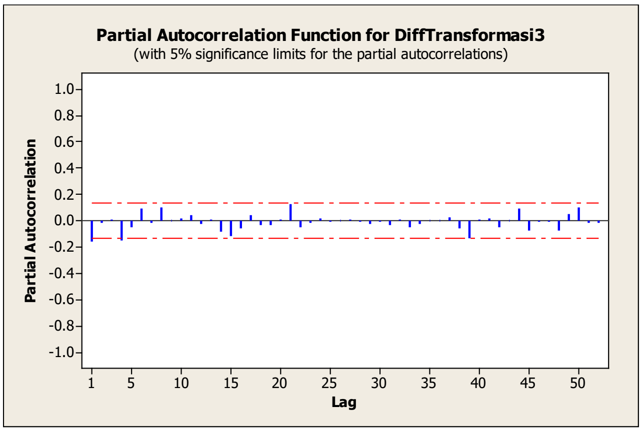

Figure 5 Partial Autocorrelation plot after first differencing

The figure shows that the ACF value is interrupted at lags 1, and 4 so that the temporary model has the order of MA (4). Then the PACF value to identify the autoregressive process looks disconnected at lags 1 and 4, so it is estimated that the autoregressive process has an order of AR(4). From the ACF and PACF plots, the initial ARIMA model (4,1,4) was obtained. Although it is possible that there are other ARIMA models that are formed.

Significance test

There are 3 significant models, namely ARIMA(0,1,1), ARIMA(1,1,0), ARIMA(2,1,3). These three models have a significance test value of $|t| > T_{tablel}$where the value of $T_{tabel} = 1{,}96$ and has a $p - value > 0{,}05$. So that the three models will be tested for residual values to determine the best ARIMA model to be used.

White Noise Test

Table 2 White Noise Test Result

Model	lag	df	Q	$x^2{}_{(\alpha,K-p-q)}$	p-value	White Noise Test
ARIMA(0,1,1)	12	10	11.08223515	18.307	0.3510	White Noise
	24	22	19.16938068	33.924	0.6350	White Noise
	36	34	21.17416798	48,602	0.9580	White Noise
	48	46	33.06159916	62,829	0.9240	White Noise
ARIMA(1,1,0)	12	10	11.13448748	18.307	0.3470	White Noise
	24	22	19.26714958	33.924	0.6290	White Noise
	36	34	21.22681386	48,602	0.9570	White Noise
	48	46	33.51779833	62.829	0.9150	White Noise
ARIMA(2,1,3)	12	6	9.071876935	12.59159	0.1700	White Noise
	24	18	15.78267205	28.8693	0.6080	White Noise
	36	30	18.03029017	43.77297	0.9580	White Noise
	48	42	29.69405383	58.12404	0.9230	White Noise

Based on the table above, the results of the white noise test at lags 12, 24, 36 and 48 have a value of $Q < x^2{}_{(\alpha,K-p-q)}$ and a value of $p - value > \alpha$. This shows that all ARIMA temporary models are not correlated with residuals so that they meet the white noise assumption.

Residual Normality Test

Table 3 Experiment Residual Normality Test

Model	D	$D_{(\alpha,n)}$	Hasil Uji
ARIMA(0,1,1)	0.086	0.094073	Normal Distribution
ARIMA(1,1,0)	0.083	0.094073	Normal Distribution
ARIMA(2,1,3)	0.144	0.094073	Not Normal Distribution

The ARIMA(0,1,1) and ARIMA(1,1,0) models have a value of $D > D_{(\alpha,n)}$ which fulfills the assumption of a normal distribution.

Best Model Selection

After performing the diagnostic test, there are still two tentative models that assume white noise and normal residuals. For this reason, it is necessary to select the best model.

Table 4 Error Values

	ARIMA(0,1,1)	ARIMA (1,1,0)
MAE	94770.53	94491.42
MAPE	10.51009272	10.47776409
MSE	12335515313	12281027973
RMSE	111065.365	110819.7996

Based on the table above, it can be seen that the value of the test results on MAD, MAPE, MSE, RMSE in the ARIMA(1,1,0) model has a smaller value than the ARIMA(0,1,1) model. So the ARIMA model that will be used to predict the weekly gold price is the ARIMA(1,1,0).

Forecasting using ARIMA(1,1,0)

The last step is to determine the gold price forecast for the next period. At this stage, the gold price will be forecasted from January 2020 to December 2021, where the general model is:

$$\Delta X_t = \mu' + \emptyset \Delta X_{t-1} + e_t$$

$$X_t - X_{t-1} = \mu' + \emptyset(X_{t-1} - \emptyset X_{t-2}) + e_t$$

$$X_t = X_{t-1} + \mu' + \emptyset(X_{t-1} - \emptyset X_{t-2}) + e_t$$

Where the value of $\mu' = 0.09265$ and the value of $\emptyset = -0.1578$. So that the ARIMA (1,1,0) model is obtained:

$$X_t = X_{t-1} + \ 0.09265 \ + -0.1578(X_{t-1} - \emptyset X_{t-2}) + e_t$$

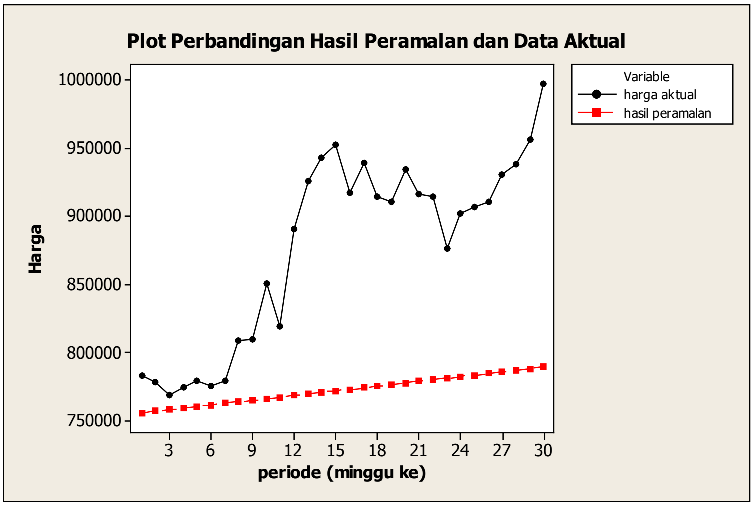

Figure 6 Comparison of Forecasting Gold Prices and Actual Gold Prices

4 Conclusion

Based on research conducted on the price of gold, it can be concluded that:ARIMA is a time series data analysis model that can predict future data. Weekly gold price forecasting in January 2016 - December 2019 the right model to forecast gold prices is ARIMA(1,1,0), with the following equation:

$$X_t = X_{t-1} + \ 0.09265 \ + -0.1578(X_{t-1} - \emptyset X_{t-2}) + e_t.$$

ARIMA value (1,1,0) with error values at MAE= 94491.42, MAPE= 10.47776409, MSE=12281027973, RMSE=110819.7996.

Based on the results of PT Antam's weekly gold price forecast from 2020 to 2021, the prediction of gold prices will increase. So if investors want to make a purchase, it is recommended that they start from 2020 or the beginning of the 2021 period. Meanwhile, old investors who want to sell gold are advised to make a sale at the end of 2021 because the price of gold has increased to Rp. 878,765.00.

REFERENCES

[1] L. Putri, and S. Soekarno, "The Analysis of Risk and Return Comparation Between Silver, Gold, JKSE, and Property in Pasar Kemis, Sukamatri, Kutajaya, Gelam Jaya and Kutabaru Wards in Kabupaten Tangerang for Long Term Investment," Journal of Business and Management, vol 1, No 3, pp. 190-195. 2012

[2] A. Halim, *Analisis Investasi,* 2nd ed., PT Salemba Emban Patria, Jakarta, p. 4-5, 2005.

[3] Jogiyanto, *Teori Portofolio dan Analisis Investasi*, BPFE, Yogyakarta, p. 285-302, 2007

[4] A. Halim, *Analisis Investasi dan Aplikasinya, dalam Aset Keuangan dan Aset Riil*, Salemba Empat, Jakarta, p.13-14, 2015