American Journal of Sciences and Engineering Research
E-ISSN -2348 – 703X, Volume 5, Issue 2, 2022

Time Series Modeling and Forecasting of Gold Prices on International Financial Markets

Chaku, S. E.[1], Gabriel, F. G.[2], Abdulrazaq A. A.[3], Aliyu, S. A.[4], Adehi M. U.[5], Timnan, B. N.[6]
[1,3,4,5]*Department of Statistics, Nasarawa State University, Keffi.*
[2]*Department of Accounting, University of Calabar, Nigeria.*
[6]*Department of Economics, University of Jos, Nigeria.*

ABSTRACT: *Application of SARIMA model in modeling and forecasting average monthly gold prices was carried out in this study. Data on gold from January 2015 to December 2020 was obtained. Monthly adjusted close prices were used for the analysis. The gold price data was stationary after first difference (D = -3.8426, P = 0.02183< 0.05). SARIMA(0,0,0)(0,1,1)[12] was identified as the best model that fit the gold price data with minimum AIC and BIC. Forecast of gold prices from January, 2021 to December 2025 was obtained. Forecast shows a rise and fall of the average monthly gold price over the forecast period (2021-2025). The Forecast values were tested against actual values for January, 2021 to June, 2021. There was no significant difference between the actual gold prices against predicted values (t = 2.102, P = 0.07191< 0.05). Prospective investors should consider gold in their portfolios as a store of value and a diversification tool and cautious of the price fluctuation predicted in this study.*

Key Words: *SARIMA Model, Gold, Seasonality, Forecast.*

I. Introduction

1.1 Background to the Study

Time series models has been commonly used in a broad range of scientific applications, including gold price forecasting; however, applications for forecasting has been limited (Ashok & Vijay, 2010). Some of the major advantages of time series models include their systematic search capability for identification, estimation, and diagnostic checking (Mishra & Desai, 2005). Time series models, like the Autoregressive Integrated Moving Average (ARIMA), effectively consider serial linear correlation among observations, whereas Seasonal Autoregressive Integrated Moving Average (SARIMA) models can satisfactorily describe time series that exhibit non-stationary behaviors both within and across seasons (Box & Jenkins, 1976).

A growing interest in gold as an investment has prompted conducting this research focused on modeling and forecasting gold prices. In this study, the researcher takes the monthly gold price data as the research object and the data between January, 2015 – December, 2020 are regarded as the sequence to fit the seasonal model. The gold price data from January to June in 2021 will be used for comparison with the forecast obtained using the fitted model.

1.2 Research Questions

This study answers the following questions:

1. What is the pattern of the average monthly gold price data?
2. What SARIMA model best fits the average monthly gold price data?

 Received-12-12-2021, Accepted- 15-01-2022

3. What is the pattern of future average monthly gold price data?

1.3 Aim and Objectives of the Study

The study aims to fit a Seasonal Autoregressive Moving Average (SARIMA) model to monthly average gold prices.

The specific objectives are to:

(i). identify patterns (seasonality ad trend) in the average monthly gold price data.

(ii). identify the best SARIMA model that fits the average monthly gold price data.

(iii). Predict future average monthly gold price using the best-estimated SARIMA model.

1.4 Scope of the Study

The study is concerned with modeling and forecasting average monthly gold prices. The study covered the period from January - December 2015 to 2020 (72 months).

II. RESEARCH METHODOLOGY

2.1 Research Design

The empirical research design was adopted in this study. With the empirical design, the researcher collects data on the topic of investigation, analyzes, and drawsa conclusion based on the results of the analysis.

2.2 Source of Data

The data used for the analysis of this study was sourced from Yahoo Finance data portal, (https://finance.yahoo.com/). The dataset contains the daily opening and closing prices of gold from January, 2015 to December, 2020.

2.3 Test of Seasonality

The presence of a seasonal effect in a series is quite obvious and the seasonal periods are easy to find (e.g., 4 for quarterly data, 12 for monthly data, etc.). Seasonality can be visually identified in the series as a pattern that repeats every *k* element.

Seasonality in time series can be identified from the time plot of the entire series by regularly spaced peaks and troughs which have a consistent direction and approximately the same magnitude every period/year, relative to the trend. In some cases, the presence of a seasonal effect in a series is not quite obvious and, therefore, testing is required to confirm the presence of the seasonal effect in a series (Box & Jenkins, 1976).

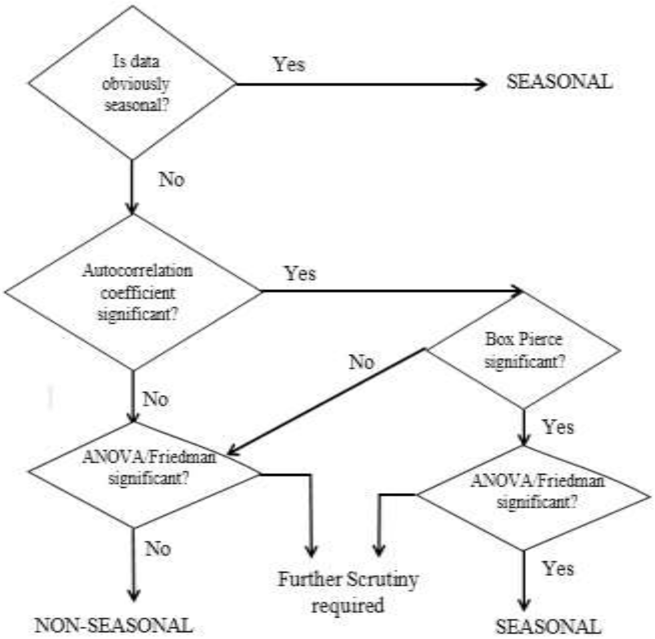

Fig 2.1: Flowchart for Identifying seasonal data

2.4 Modeling Procedure

The Box Jenkin methodology was employed to identify the best seasonal ARIMA model for the average monthly gold price data. By the Box Jenkin, when fitting an ARIMA model to a set of time-series data, the following procedure provides a useful general approach:

1. Plot the data. Identify any unusual observations.
2. If necessary, transform the data (using a Box-Cox transformation) to stabilize the variance.
3. If the data are non-stationary: take the first differences of the data until the data are stationary.
4. Examine the ACF/PACF: Is an AR(pp) or MA(qq) model appropriate?
5. Try your chosen model(s), and use the AICc to search for a better model.
6. Check the residuals from your chosen model by plotting the ACF of the residuals, and doing multiple tests of the residuals. If they do not look like white noise, try a modified model.
7. Once the residuals look like white noise, calculate forecasts.

2.5 Model Selection Criteria

In choosing the model that best describes a time series data, attention is given to the RMSE, AIC, BIC. Smaller values indicate a better model.

2-log Likelihood: The -2 log-likelihood is the most basic measure for model selection. It is the basis for the AIC and BIC. $L = -Nln(\sigma^2{}_a) - \frac{SSQ'}{2\sigma^2{}_a} - \frac{2Nln(2\pi)}{2}$

AIC: Akaike's Information Criterion (AIC) adjusts the -2 Restricted Log-Likelihood by twice the number of parameters in the model. $AIC = -2L + 2N_p$

BIC: The Bayesian information criterion (BIC) is a measure for selecting and comparing models based on the -2 log-likelihood. Smaller values indicate better models. The BIC also penalizes over-parametrized models, but more strictly than the AIC because the BIC accounts for the size of the dataset as well as the size of the model. $BIC = -2L + \ln(N)N_p$

Where; N = Total number of observations, L= -2 log likelihood,$\sigma^2{}_a$=variance of residuals

N_p= Number of parameters $\left(N_p = p + q + d + P + Q + D + m\right)$

III. DATA PRESENTATION AND ANALYSIS

3.1 Data Presentation

Table 3.1: Average price (USD) of gold from Jan, 2015 – Dec, 2020

Month/Year	2015	2016	2017	2018	2019	2020
Jan	1,285.42	1,088.30	1,181.50	1,319.00	1,286.20	1,552.10
Feb	1,204.40	1,172.30	1,226.30	1,325.70	1,311.70	1,566.90
Mar	1,178.00	1,256.80	1,217.80	1,321.00	1,291.30	1,641.90
Apr	1,203.50	1,237.00	1,259.90	1,337.80	1,294.00	1,675.80
May	1,190.10	1,277.50	1,231.70	1,311.70	1,282.70	1,703.90
Jun	1,180.70	1,248.70	1,276.00	1,296.70	1,330.60	1,712.70
Jul	1,160.60	1,348.60	1,219.00	1,247.60	1,401.90	1,798.30
Aug	1,100.10	1,347.40	1,269.90	1,209.80	1,479.20	1,989.80
Sep	1,146.30	1,328.50	1,332.90	1,196.60	1,510.50	1,941.90
Oct	1,098.90	1,261.30	1,279.00	1,196.70	1,497.20	1,902.80
Nov	1,071.60	1,278.80	1,276.20	1,219.60	1,475.80	1,894.60
Dec	1,098.90	1,167.60	1,255.60	1,241.10	1,467.60	1,839.20

Source: Yahoo Finance (https://finance.yahoo.com/)

3.2 Data Analysis and Results

3.2.1 Descriptive Statistics

Table 3.2: Descriptive Statistics

Statistic	Estimate	Statistic	Estimate
N	72	Range	918.2
Missing	0	Minimum	1071.6
Mean	1,347.6	Maximum	1989.8
Std. error mean	25.6	Coefficient of variation	16.13
Median	1278.9	Skewness	1.47
Mode*	1311.7	Kurtosis	1.48
Sum	97,029.0	Sum of squares	134112743.9
Standard deviation	217.3	1st Quartile	1211.8
Variance	47239.2	3rd Quartile	1388.6

**More than one mode exists, only the first is reported*

Table 3.2 present the descriptive statistics of the average monthly price of gold (USD) from January 2015 to December 2020. There are 72 observations and no missing values. The average price of gold over the study period is $1,347.6. The Coefficient of variation of 16.13 means there is 16.13% variation in the dataset. Skewness and Kurtosis (1.47 and 1.48 respectively) show the data is not skewed and the distribution is not peaked. The minimum and maximum monthly average prices of gold over the study period are $1,071.6 and $1,989.8.

3.2.2 Time-plot & Model Identification

A Seasonal Autoregressive Integrated Moving Average (SARIMA) model of the form $SARIMA(p,d,q)\times(P,D,Q)_{[s]}$ was fitted for the average monthly price of gold from 2015-2020. The model is given by:

$(1-B)^d(1-B^s)^D\varphi(B)\Phi(B)X_t= \vartheta(B)\Theta(B)\varepsilon_t$

Where,

P and p = orders of seasonal and non-seasonal Autoregression.

Q and q = orders of seasonal and non-seasonal Moving Average (MA) respectively.

D and d = seasonal and non-seasonal difference.

m = order of seasonality.

X_t is the time series value at time t and φ, ϑ, Φ and Θ, are polynomials of the order p, q, P, and Q respectively. B is the backward shift operator and the white noise process is denoted by ε_t.

$s = 12$since the gold price series is made of monthly observations. therefore;

$(1-B)^d(1-B^{12})^D\varphi(B)\Phi(B)X_t= \vartheta(B)\Theta(B)\varepsilon_t$

To get an insight from the data, it is decomposed to disclose different components of the time series into raw data, seasonality, trend, and random error.

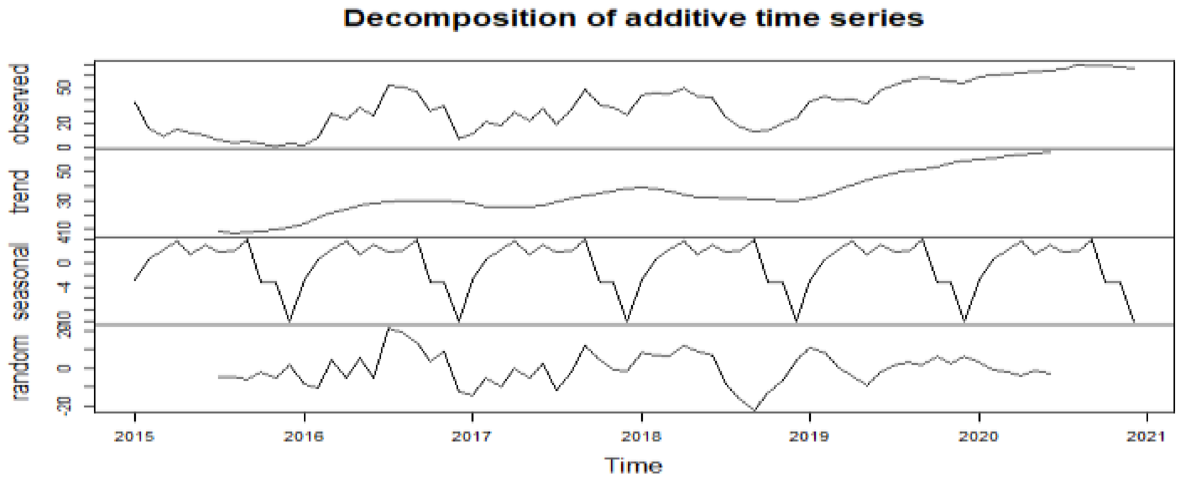

Fig. 3.1: Decomposition of the gold price (2015-2020)

Figure 3.1 present the decomposition of the gold price data. It breaks the series into four components. The observed section shows the original gold price data. The trend component shows there was a continuous increase (upward trend) of the gold price over the study period (2015-2020). The price of gold is at its peak in 2020 than all the past years in the study period. The seasonal part of the decomposition shows a seasonal pattern (a regular rise and fall every year). The random component of the decomposition shows that the series is not a stationary process.

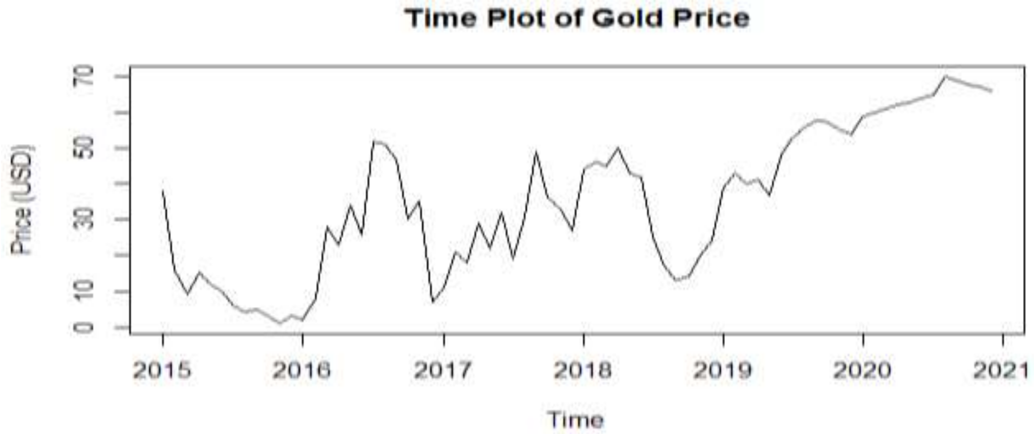

Fig. 3.2: Time plot of gold price data (2015-2020)

Figure 3.2 shows the time plot of the gold price series. The plot shows the series is not stationary in the mean, and since is a seasonal record of monthly data, taking seasonal difference (length 12) seems appropriate.

Table 3.3: Augmented Dickey-Fuller Test

Dataset	Dickey-Fuller statistic	Lag order	p-value	Remark
Original Data	-3.3414	4	0.07208	Not Stationary
1st Differenced Data	-3.8426	4	0.02183	Stationary

Source: R-output

Table 3.3 present the Augmented Dickey-Fuller (ADF) Test of stationarity. It shows that the gold price series is stationary after the first difference.

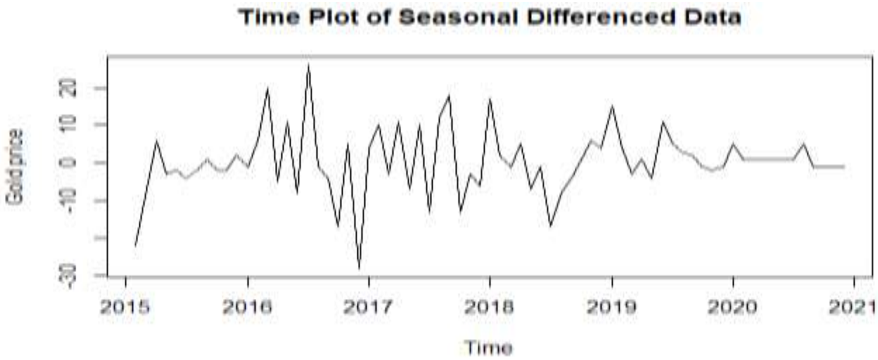

Fig. 3.3: Time Plot of seasonally differenced data

Figure 3.3 present the time plot of seasonally differenced data. It shows the 1st differenced gold price data is stationary.

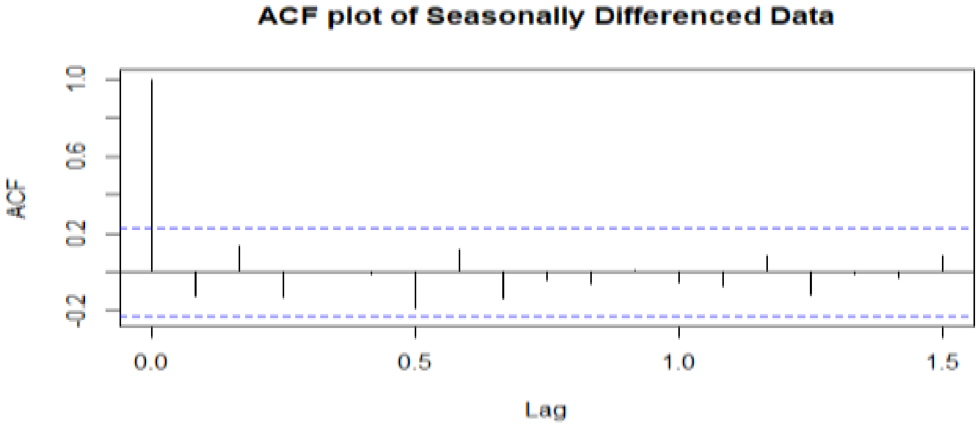

Fig. 3.4: ACF of seasonally differenced data

Figure 3.4 presents the ACF (Autocorrelation Function) plot of the seasonally differenced data. It shows there is one significant spike at the seasonal lag (lag 0). This is suggestive of a seasonal moving average SMA(1) model. Since there are no significant spikes at the non-seasonal lags, this suggests a non-seasonal moving MA(0) model. This sets q=0 and Q=1 in the SARIMA(p,d,q)(P,D,Q)[12] model.

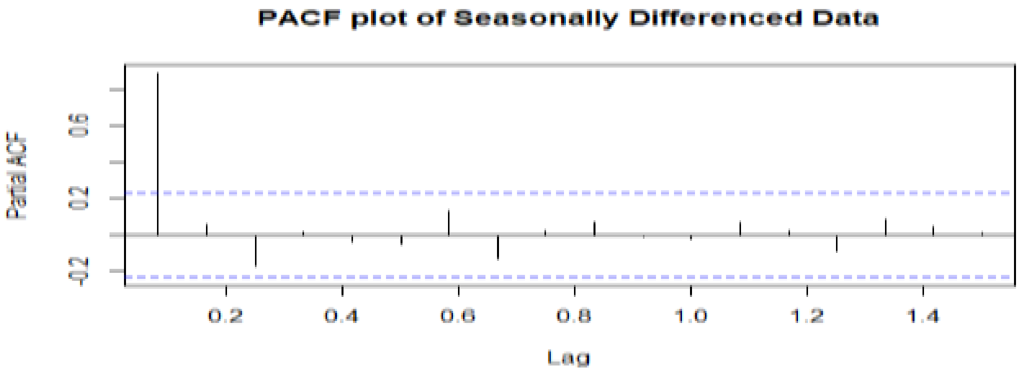

Fig. 3.5: PACF of seasonally differenced data

Figure 3.5 present the PACF (Partial Autocorrelation Function) of the seasonally differenced data. There is one significant spike at the seasonal lag. This is suggestive of the seasonal SAR(1) model. There are no significant spikes at the non-seasonal lags, which is suggestive of the non-seasonal AR(0) model. The result of the PACF sets p=0 and P=1 in SARIMA(p,d,q)(P,D,Q)[12] model.

Hence, this initial analysis suggests that a possible model for the data is of the form:

Non-seasonal component (p,d,q); as p = 0, d = 0, q = 0

Seasonal component (P,D,Q) with P = 1, D = 1 and Q = 1.
The seasonal length is twelve (12) months thus the initial model is SARIMA(0,0,0)(1,1,1)[12].
The model SARIMA(0,0,0)(1,1,1)[12], along with some variations of it, are fitted and their accuracy measures - AIC, and BIC computed as seen in the table below.

Table 3.4: Different combinations of SARIMA models by several metrics

S/N	SARIMA Model SARMA(p,d,q)(P,D,Q)[12]	AIC	BIC
1	(0,0,0)(1,1,1)[12]	453.7479*	459.9805*
2	(1,0,0)(0,1,1)[12]	453.3397	459.5723
3	(0,0,1)(0,1,1)[12]	453.4945	459.7271
4	(0,0,0)(0,1,1)[12]	452.1452**	456.3003**
5	(1,0,1)(0,1,1)[12]	454.5702	462.8803
6	(1,1,0)(0,1,1)[12]	463.7465	469.9278
7	(1,1,0)(1,1,1)[12]	465.4226	473.6644
8	(1,1,0)(1,1,0)[12]	474.643	480.8244
9	(1,0,0)(0,1,0)[12]	479.4735	483.6285
10	(0,1,1)(0,1,1)[12]	452.2707	458.4521
11	(1,1,1)(0,1,1)[12]	453.5846	461.8264
12	(1,0,1)(1,0,1)[12]	520.3513	533.9274

***The minimum of AIC, and BIC.*
**Initial model*
Source: R-output

Table 3.4 presents twelve (12) different SARIMA models obtained using the "ARIMA()" function in R-software and their accuracy measures (AIC and BIC). Of these models, SARIMA(0,0,0)(0,1,1)[12] is identified as the best model with the minimum of AIC and BIC.

3.3 Estimation of Model Parameters

Table 3.5: Parameter Estimate for SARIMA(0,0,0)(0,1,1)[12]

Coefficients	Estimate	Std. Error	z value	Pr(>|z|)
SMA(1)	-1.00000	0.22419	-4.4604	0.000008***

Sigma ^2 estimated as 81.16 Log likelihood= -224.07

AIC=**452.1452** BIC=456.3003 RMSE=8.212555

*Sig. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1*

Table 3.5 present the estimate and test of significance of the parameters of SARIMA(0,0,0)(0,1,1)[12]. The seasonal moving average parameter is significant (P<0.05). The model can be written as: $(1\text{-}B^{12})X_t = \Theta(B)\varepsilon_t$
Substituting the model parameter gives rise to: $(1 - B^{12})X_t = (1 - 1.0B^{12})\varepsilon_t$

4.2.4 Model Diagnostic Check

Table 3.6: Shapiro-Wilk normality test of residuals

Data	Wilk	P-value
residuals(fit4); SARIMA(0,0,0)(0,1,1)[12].	0.96539	0.04733
Null Hypothesis: Distribution is not normal		

Source: R-output

Table 3.6 present the Shapiro-Wilk normality test for residuals from SARIMA(0,0,0)(0,1,1)[12]. The closer the Wilk statistic (W) is to 1, the better the distribution is normally distributed. The P-value indicates the rejection of the null hypothesis that the distribution of the residuals is approximately normal. The P-value shows the residuals are normally distributed (P<0.05).

Table 3.7: Ljung-Box test of residuals

Data	Q*	Df	Model df	P-value
residuals(fit4); SARIMA(0,0,0)(0,1,1)[12].	3.2349	13	1	0.9969
Null Hypothesis: The model is adequate				

Source: R-output

From Table 3.7, the P-value indicates there is no evidence that the residuals are independent. This further confirms that the SARIMA(0,0,0)(0,1,1)[12] is adequate. If the P-value was below about 0.05, there would be some cause for concern: it would imply that the terms in the ACF are too large to be white noise.

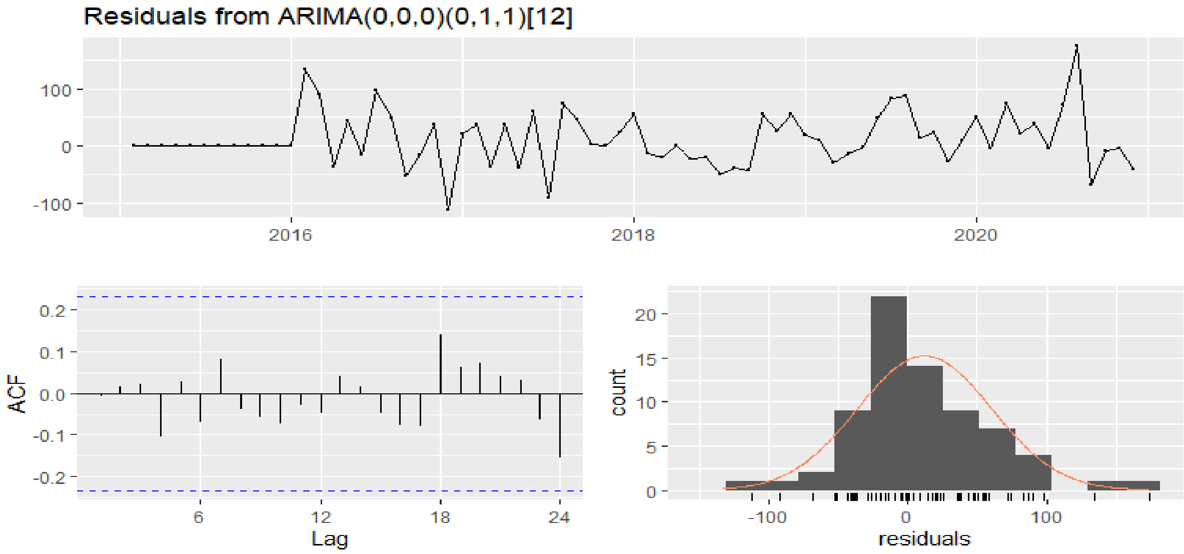

Fig. 3.6: Plot residuals plot

Figure 3.6 present the graphical diagnostic of residuals from SARIMA(0,0,0)(0,1,1)[12]. The time plot shows that the residuals have no definite pattern. The ACF plot shows that all the points in the correlogram are within the threshold limits. The histogram shows that the distribution of the residuals is bell-shaped which indicates a normal distribution. These unveiled that the residuals from the SARIMA model fit the data and can be used for forecasting.

3.5 Forecast of Gold Price

Table 3.8: Forecast of Gold Price from January 2021 to December 2025

Month/Year	Pont	lower limit	Upper Limit
Jan-21	1,570.57	1,505.36	1,635.78
Feb-21	1,257.84	1,192.63	1,323.05
Mar-21	1,302.94	1,237.73	1,368.15
Apr-21	1,269.65	1,204.44	1,334.86
May-21	1,090.04	1,024.83	1,155.25
Jun-21	1,202.75	1,137.54	1,267.96
Jul-21	1,422.92	1,357.71	1,488.13

Aug-21	1,763.77	1,698.56	1,828.98
Sep-21	1,085.76	1,020.55	1,150.97
Oct-21	752.90	687.69	818.11
Nov-21	1,034.73	969.52	1,099.94
Dec-21	815.83	750.62	881.04
Jan-22	1,613.24	1,548.03	1,678.45
Feb-22	1,300.68	1,235.47	1,365.89
Mar-22	1,345.77	1,280.56	1,410.98
Apr-22	1,312.49	1,247.28	1,377.70
May-22	1,132.88	1,067.67	1,198.09
Jun-22	1,245.59	1,180.38	1,310.80
Jul-22	1,465.76	1,400.55	1,530.97
Aug-22	1,806.61	1,741.40	1,871.82
Sep-22	1,128.60	1,063.39	1,193.81
Oct-22	795.74	730.53	860.95
Nov-22	1,077.57	1,012.36	1,142.78
Dec-22	858.67	793.46	923.88
Jan-23	1,654.34	1,589.13	1,719.55
Feb-23	1,341.94	1,276.73	1,407.15
Mar-23	1,387.04	1,321.83	1,452.25
Apr-23	1,353.75	1,288.54	1,418.96
May-23	1,174.14	1,108.93	1,239.35
Jun-23	1,286.85	1,221.64	1,352.06
Jul-23	1,507.02	1,441.81	1,572.23
Aug-23	1,847.87	1,782.66	1,913.08
Sep-23	1,169.86	1,104.65	1,235.07
Oct-23	837.00	771.79	902.21
Nov-23	1,118.83	1,053.62	1,184.04
Dec-23	899.93	834.72	965.14
Jan-24	1,694.04	1,628.83	1,759.25
Feb-24	1,381.79	1,316.58	1,447.00
Mar-24	1,426.88	1,361.67	1,492.09
Apr-24	1,393.60	1,328.39	1,458.81
May-24	1,213.99	1,148.78	1,279.20
Jun-24	1,326.70	1,261.49	1,391.91
Jul-24	1,546.87	1,481.66	1,612.08
Aug-24	1,887.72	1,822.51	1,952.93
Sep-24	1,209.71	1,144.50	1,274.92
Oct-24	876.84	811.63	942.05
Nov-24	1,158.68	1,093.47	1,223.89
Dec-24	939.78	874.57	1,004.99
Jan-25	1,732.48	1,667.27	1,797.69
Feb-25	1,420.36	1,355.15	1,485.57
Mar-25	1,465.45	1,400.24	1,530.66
Apr-25	1,432.17	1,366.96	1,497.38
May-25	1,252.56	1,187.35	1,317.77
Jun-25	1,365.27	1,300.06	1,430.48

Jul-25	1,585.44	1,520.23	1,650.65
Aug-25	1,926.29	1,861.08	1,991.50
Sep-25	1,248.28	1,183.07	1,313.49
Oct-25	915.41	850.20	980.62
Nov-25	1,197.25	1,132.04	1,262.46
Dec-25	1,248.28	1,183.07	1,313.49

Source: SARIMA(0,0,0)(1,1,1)[12] from R-output

Table 3.8 presents five (5) years' forecast of average monthly gold prices from January 2021 to December 2025.

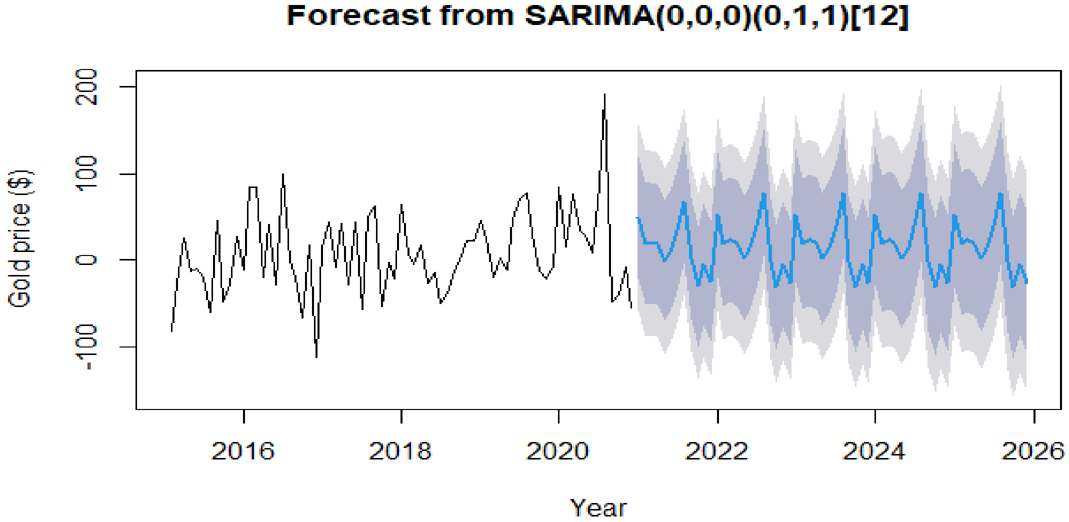

Fig. 3.7: Forecast plot of gold prices

Figure 3.7 shows the gold price series and forecast from SARIMA(0,0,0)(0,1,1)[12]. It shows a regular rise and fall of the average monthly gold price over the forecast period (2021-2025).

Table 3.9: T-test of actual gold prices against predicted values (Jan-June, 2021)

Month	Actual price ($)	Predicted price ($)	Mean Difference ($)	T-Stat	P-value
January 2021	1,866.58	1,570.57			
February 2021	1,805.53	1,257.84	523.73	2.102	0.07191
March 2021	1,719.52	1,302.94			
April 2021	1,759.76	1,269.65			
May 2021	1,850.65	1,090.04			
June 2021	1,834.12	1,202.75			

Source: R-output

Table 3.9 present the t-test of actual gold prices against predicted values from January-June, 2021. The test shows there is no significant difference between the actual and the predicted gold prices.

IV. Discussion of Findings and Conclusion

4.1 Discussion of Findings

The average price of gold over the study period (2015-2020) is $1,347.6. Decomposition of the gold price data shows there was a continuous increase (upward trend) of the gold price over the study period (2015-2020). The price of gold is at its peak in 2020 than all the past years in the study period.

Augmented Dickey-Fuller (ADF) test shows the gold price data is stationary at first difference. This agreed with the findings of Abdullah (2012),Davis, Dedu, &Bonye (2014), Guha&Bandyopadhyay (2016), Tripathy (2017), and Ligita et al. (2018).

Initial analysis presents SARIMA(0,0,0)(1,1,1)[12] as the suggested model for the data. Comparison with several models shows that SARIMA(0,0,0)(0,1,1)[12] is the best model that fits the gold price data with the minimum of AIC and BIC.

Shapiro-Wilk test shows that residuals form SARIMA(0,0,0)(0,1,1)[12] are normally distributed ($P<0.05$). Ljung-Box test of residuals indicates there is no evidence that the residuals are dependent ($P>0.05$).

The forecast shows a regular rise and fall of the average monthly gold price over the forecast period (2021-2025).

T-test for comparison of actual gold prices against predicted values from January-June, 2021shows there is no significant difference between the actual and the predicted gold prices ($P>0.05$).

4.2 Conclusion

Gold price data shows there was a continuous increase (upward trend) of the gold price over the study period (2015-2020). The price of gold was at its peak in 2020 than all the past years in the study period. Augmented Dickey-Fuller (ADF) test shows the gold price data is stationary at first difference. SARIMA (0,0,0)(0,1,1)[12] is the best model that fits the gold price data with the minimum AIC and BIC values. The forecast shows a regular rise and fall of the average monthly gold price over the forecast period (2021-2025). There was no significant difference between the actual gold prices against predicted values from January-June, 2021.

V. REFERENCES

1. Abdullah L. (2012):ARIMA Model for Gold Bullion Coin Selling Prices Forecasting. International Journal of Advances in Applied Sciences 1(4):153-158
2. Ashok K. M. &Vijay P. S. (2010): A Review of Drought Concepts*J. Hydrol*. 391, pp. 202-216
3. Box G.E.P., and JenkinsG.M. (1976): JenkinsTime Series Analysis: Forecasting and Control, John Wiley & Sons. Holden-Day, San Francisco.
4. Davis R, Dedu V K, Bonye F. (2014): Modeling and Forecasting of Gold Prices on Financial Markets. *American International Journal of Contemporary Research*, 4:107-113
5. Guha B, Bandyopadhyay G. (2016): Gold Price Forecasting Using ARIMA Model. *Journal of Advanced Management Science*, 4(2):117-121
6. Latifa G.& Sara F. (2020): Modeling and Forecasting Gold Prices. *In: Journal of Economics and Business*, 3(4), 1708-1729.
7. LigitaG., Remeikiene, R., Sadeckas, A., &Ginevicius, R. (2018). Gold Investment Incentives: An Empirical Identification of the Main Gold Price Determinants and Prognostication of Gold Price Future Trends. *Economics and Sociology*, *11*(3), 248-264. doi:10.14254/2071-789X.2018/11-3/15
8. Mishra A. K. &Desai V. R. (2005): Spatial and temporal drought analysis in the Kansabati River Basin, *India Int. J. River Basin Manage*, 3 (1), pp. 31-41
9. Tripathy N. (2017): Forecasting Gold Price with Auto-Regressive Integrated Moving Average Model. *International Journal of Economics and Financial Issues.* 7:324-329

.