2024 SICASH
SLIIT INTERNATIONAL CONFERENCE ON ADVANCEMENTS IN SCIENCES & HUMANITIES

Development of Time Series Model to Predict Daily Gold Price

Gayashan, W. A. K.*[1], Dayarathna, A. K. G.[2], Rajakaruna, R. W. M. A. P.[3], Perera, T. J. N.[4], Peiris, T. S. G.[5]

[1,2,3,4,5] *Sri Lanka Institute of Information Technology, Malabe, Sri Lanka*

Email address of the corresponding author -*kavindu0813@gmail.com

Abstract

Gold is ancient and one of the most precious and popular commodities in the world. Gold price forecasting is critical in financial decision-making, providing valuable information for investors in the gold market, sellers of gold items and stakeholders. Not much studies have been carried out in to forecast daily gold prices of Sri Lanka. The aim of this paper is to forecast the daily gold price rate (Rupees/troy ounce) using data from 2nd January 2018 to 14th June 2024 published by the Central Bank of Sri Lanka. The best fitted model was identified as ARIMA (1,1,1) + ARCH (2). The model was trained using data from 2nd January 2018 to 31st May 2024 and validated using data from the 3rd of June 2024 to 14th of June 2024. The model was statistically tested using standard statistical procedure and errors were found as white noise. The Mean Absolute Percentage Error (MAPE) for the training data set and validation data set were 0.748% and 1.002% respectively. The validation confirmed that the ARIMA (1,1,1) + ARCH (2) model effectively captures the dynamics of gold price movements, offering robust predictive power. These results indicate that the model is highly accurate and reliable for forecasting, making it a valuable tool for financial institutions and investors aiming to predict market trends and make informed investment decisions.

Keywords: gold price rates; ARIMA models; Forecasting

Introduction

Gold, a precious metal known for its lustrous yellow appearance, has been integral to human history and economics for thousands of years(Sara Farhat, Modeling and Forecasting Gold Prices, 2020).It is celebrated not only for its beauty and use in jewellery but also for its intrinsic value and significance in financial systems. Verifiably, gold has been utilized as currency, a store of wealth, and a symbol of power. Its rarity and unique physical properties, such as resistance to tarnish and corrosion, have made it an ideal choice for coins, bullion, and monetary reserves. Central banks around the world hold substantial amounts of gold in their foreign exchange reserves, highlighting its enduring role in global finance. In modern economies, gold is traded on international or universal markets, with its price determined by supply and demand dynamics. Investors often turn to gold as a safeguard against inflation and economic uncertainty, viewing it as a stable asset during market volatility. The price of gold is often measured per troy ounce, and it is influenced by various factors, including geopolitical events, interest rates, currency movements, and broader economic conditions

https://doi.org/10.54389/WYML9575

(Hayes, (July 09, 2024). Gold vs. Stocks and Bonds. *Has Gold Been a Good Investment Over the Long Term*.).

The world gold market has undergone significant fluctuations from 2018 to 2024, reflecting a variety of economic, political, and social influences. This period has been marked by notable events such as trade tensions between major economies, the global COVID-19 pandemic, fluctuating interest rates, and geopolitical instability(Wee Chian Koh, J. B. (July 27, 2020). Gold price. *Gold shines bright throughout the COVID-19 crisis*). These factors have collectively impacted investor sentiment and, consequently, the daily gold prices. This research aims to provide a comprehensive and more detailed analysis with forecasted data (Using Time Series analysis) of these price movements, exploring the factors that have influenced the gold market over the past six years. By understanding these dynamics, we can gain valuable insights into the economic forces at play and the outlook for gold as a vital budgetary resource.

Although factors such as inflation, demand, interest rates, and geopolitical events can significantly influence gold prices, these external variables were not considered in this analysis. The objective of this study is to develop a model based solely on past gold price data.

Materials and Methodology

Secondary Data

The daily gold price rate (Rupees/troy ounce) from 2nd January 2018 to 14th June 2024, was obtained from the website of the Central Bank of Sri Lanka (Central Bank Sri Lanka.(2018). *Gold Price (in LKR)*. Retrieved from cbsl.gov). The dataset excludes holidays and weekends gold prices. The dataset from 2nd January 2018 to 31st May 2024 was used to train the models and the balance data was used to validate model. The statistical analysis was performed using EViews 12 and Minitab software.

Methodology

An Auto-Regressive Integrated Moving Average (ARMA) of order p and q is represented by the equation (1).

$$Y_t = \mu + \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \dots + \varphi_p Y_{t-p} + e_t - \theta_1 e_1 - \theta_2 e_2 - \dots\dots - \theta_q e_{t-q} \quad (1)$$

Where p and q are the order of the autoregressive part and moving average part respectively. $\{e_t\}$ is the error series and it is white noise (Robert & Shumway, (2017). *Time Series Analysis and Its Applications.* Springer Cham). ARMA models are well-suited for analyzing and forecasting stationary time series data. If non-stationary behavior was detected, the differencing technique is used to make series stationary. It is denoted by ARIMA (p, d, q) where d is the different step used to achieve stationarity and p, q are indicated the autoregressive term and moving average term respectively.

However, most financial time series have more volatility points; thus, the error is heteroscedasticity. That $V(e_t) = \sigma_t^2$.In that case to estimate variance ARCH and GARCH models have been recommended. Thus, the ARCH model was designed to capture the volatility clustering observed in the gold price data. The ARCH model of order q is represented by equation (2).

$$\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \alpha_2 \varepsilon_{t-2}^2 + \dots\dots + \alpha_q \varepsilon_{t-q}^2 \quad (2)$$

The order of the ARCH model was determined based on the ACF/PACF plots of residual squared. Generally, the order of ARCH model does not consider more than 2.

Results and Discussion

Temporal Variability

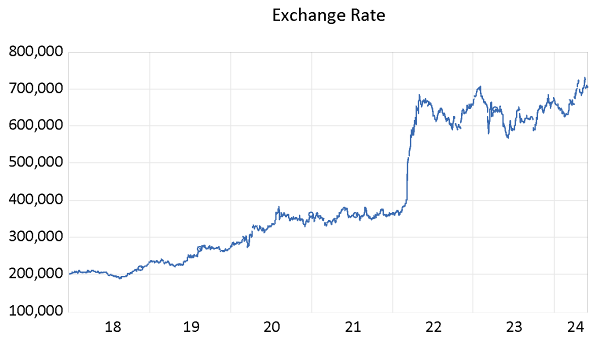


Figure 1. Temporal Variability of Original Series

Figure 1 illustrates the temporal variability of daily gold prices from 2018 to 2024. It indicates an increasing linear trend from the beginning until 2021, and after 2022, the gold price rate fluctuates. The daily gold prices vary from the minimum of Rs.187876.2(8/16/2018) to the maximum of Rs.729715.5(5/20/2024) with the mean of Rs.410933.1 and the standard deviation of 175644.0. The dataset represents a slight positive skewness with the value of 0.412047 and the significance of Jarque-Bera test confirms (178.3892, p=0.00) that the gold price significantly has deviated from normality.

Model Selection

Original Series

Date: 06/15/24 Time: 15:35
Sample: 1/02/2018 5/31/2024
Included observations: 1546

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.998	0.998	1543.2	0.000
		2	0.996	0.014	3081.7	0.000
		3	0.994	-0.041	4615.1	0.000
		4	0.992	-0.006	6143.4	0.000
		5	0.990	0.005	7666.6	0.000
		6	0.988	-0.050	9184.1	0.000
		7	0.986	0.006	10696.	0.000
		8	0.984	0.005	12203.	0.000
		9	0.982	0.007	13704.	0.000
		10	0.980	-0.015	15199.	0.000
		11	0.978	0.008	16689.	0.000
		12	0.975	-0.002	18174.	0.000
		13	0.973	-0.010	19653.	0.000
		14	0.971	-0.008	21126.	0.000
		15	0.969	0.024	22594.	0.000
		16	0.967	0.007	24057.	0.000
		17	0.965	-0.005	25514.	0.000
		18	0.963	0.007	26967.	0.000
		19	0.961	0.007	28414.	0.000
		20	0.959	-0.040	29856.	0.000
		21	0.957	-0.008	31292.	0.000
		22	0.954	0.002	32722.	0.000
		23	0.952	-0.028	34147.	0.000
		24	0.950	-0.012	35566.	0.000
		25	0.948	-0.015	36979.	0.000
		26	0.945	-0.038	38385.	0.000

Figure 2. ACF and PACF of the Original Series

Figure:2 depicts the ACF of the original series. The Augmented Dickey Fuller (ADF) test statistic for the original series, is not significant (t = -0.1274, p= 0.9446) confirming the original is not stationary.

Therefore, to check whether the series is stationary, the 1st difference series was taken, and it was found test statistics is significant (t = -25.333, P=0.9446). Thus, to identify possible ARIMA models the plot of ACF and PACF of the stationary series was considered. It is shown in Fig. 3.

Stationary Series

Date: 06/15/24 Time: 15:36
Sample (adjusted): 1/03/2018 5/31/2024
Included observations: 1545 after adjustments

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	-0.023	-0.023	0.7953	0.373
		2	0.102	0.102	16.910	0.000
		3	0.044	0.049	19.919	0.000
		4	0.010	0.002	20.075	0.000
		5	0.028	0.019	21.325	0.001
		6	-0.033	-0.036	23.052	0.001
		7	-0.037	-0.044	25.142	0.001
		8	0.049	0.052	28.827	0.000
		9	0.024	0.039	29.747	0.000
		10	0.053	0.049	34.198	0.000
		11	0.036	0.031	36.202	0.000
		12	0.016	0.004	36.589	0.000
		13	0.041	0.024	39.175	0.000
		14	-0.014	-0.018	39.462	0.000
		15	-0.043	-0.049	42.289	0.000
		16	0.019	0.019	42.827	0.000
		17	-0.011	0.004	43.021	0.000
		18	-0.001	-0.004	43.022	0.001
		19	0.074	0.074	51.525	0.000
		20	0.015	0.018	51.866	0.000
		21	0.018	-0.009	52.357	0.000
		22	0.071	0.058	60.197	0.000
		23	0.024	0.026	61.072	0.000
		24	0.026	0.009	62.096	0.000
		25	0.023	0.024	62.939	0.000
		26	-0.041	-0.041	65.613	0.000

Figure 3. ACF and PACF for the 1st Difference Series

Possible Models

The comparison of the observed ACF and PACF data and that of the theoretical ACF and PACF, the following 3 models were selected. As the second auto correlation as well as second partial auto correlation is significant, the most suitable three ARIMA models are:

ARIMA (1,1,1)

ARIMA (1,1,2)

ARIMA (2,1,2)

It was found that the errors of each model are not random. Therefore, ACF and PACF of the squared residuals were taken and it was difficult to identify any order of ARCH for the given mean equations so, the order of ARCH one and two were tested for all 3 models.

Table 1. Summary of the Significance of Models

Model	Significance of ARCH model	Significance of ARCH (1)	Significance of ARCH (2)
ARIMA (1,1,1)	Significant	Significant	Significant
ARIMA (1,1,2)	Significant	Significant	Not significant
ARIMA (2,1,2)	Significant	Significant	Not significant

Table 2. Summary of the Estimated Possible Models with ARCH

	ARIMA (1,1,1) ARCH (1)	ARIMA (1,1,2) ARCH (1)	ARIMA (2,1,2) ARCH (1)	ARIMA (1,1,1) ARCH (2)	ARIMA (1,1,2) ARCH (2)	ARIMA (2,1,2) ARCH (2)
Significant coefficients	2	2	2	2	1	0
Log likelihood	-15373.68	-15371.70	-15357.77	-15266.12	-15261.69	-1530.53
Akaike (AIC)	19.92057	19.91931	19.91545	19.78254	19.78810	19.84514
Schwartz (SIC)	19.93787	19.94006	19.93969	19.80330	19.80232	19.87284
Hannan-Quinn (HQI)	19.92701	19.92703	19.92447	19.79026	19.78711	19.85545

According to Table 2, ARIMA (1,1,1) has all significant coefficients, maximum log likelihood, lowest AIC, lowest SIC and HQI values. Therefore, ARIMA (1,1,1) ARCH (2) is the best fitted model compared to other ARCH models. The best fitted model can be written as:

Mean equation,

$$Y_t = 212.34 + 0.879Y_{t-1} + e_t + 0.837e_{t-1}$$

$$(1-0.879B)Y_t = (1+0.873B)\, e_t + 212.34$$

Variance equation,

$$\sigma_t^2 = 21054345 + 0.295\varepsilon_{t-1}^2 + 0.195\varepsilon_{t-2}^2$$

Error Diagnosis

Date: 07/12/24 Time: 19:21
Sample (adjusted): 1/04/2018 5/31/2024
Q-statistic probabilities adjusted for 2 ARMA terms

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob*
		1	-0.012	-0.012	0.2275	
		2	0.007	0.007	0.3021	
		3	0.038	0.038	2.5479	0.110
		4	-0.016	-0.015	2.9532	0.228
		5	-0.023	-0.024	3.7939	0.285
		6	-0.044	-0.046	6.7806	0.148
		7	-0.021	-0.021	7.4564	0.189
		8	0.024	0.026	8.3787	0.212
		9	-0.016	-0.013	8.7925	0.268
		10	0.013	0.012	9.0381	0.339
		11	0.036	0.031	11.005	0.275
		12	0.003	0.003	11.019	0.356
		13	0.056	0.054	15.877	0.146
		14	-0.033	-0.033	17.619	0.128
		15	-0.052	-0.054	21.919	0.057
		16	0.016	0.012	22.310	0.072
		17	-0.023	-0.014	23.168	0.081
		18	0.004	0.010	23.194	0.109

*Probabilities may not be valid for this equation specification.

Figure 4. ACF and PACF of the Residuals of the Best Fitted Model

The correlogram of residuals from the best-fitted model is represented in Figure 4. The Q statistic of residuals, p values are greater than 0.05. Therefore, it can be concluded with 95% confidence that errors are identically and independently distributed.

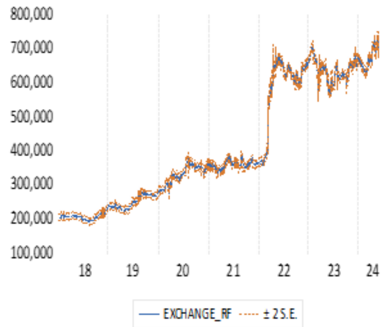
The constant variance of the error series was confirmed from the no systemic pattern of scatter plot between residuals and predicted values. Furthermore, after testing for the ARCH effect on possible ARIMA models, the best-fitting model was selected. The heteroscedasticity test for this model resulted in an ARCH p-value of 0.8852, which is greater than the significance level of 0.05. Therefore, we can conclude with 95% confidence that there is no ARCH effect present.

Heteroskedasticity Test: ARCH

F-statistic	0.020855	Prob. F(1,1541)	0.8852
Obs*R-squared	0.020882	Prob. Chi-Square(1)	0.8851

Figure 5. Heteroskedasticity Test for ARCH

The forecast values for the training data set along with ± 2SE,



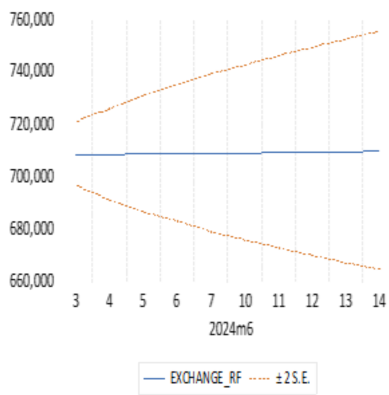
Forecast: EXCHANGE_RF	
Actual: EXCHANGE_RATE	
Forecast sample: 1/02/2018 5/31/2024	
Adjusted sample: 1/04/2018 5/31/2024	
Included observations: 1544	
Root Mean Squared Error	5474.044
Mean Absolute Error	3200.292
Mean Abs. Percent Error	0.747459
Theil Inequality Coef.	0.006122
Bias Proportion	0.000232
Variance Proportion	0.000038
Covariance Proportion	0.999687
Theil U2 Coefficient	0.996209
Symmetric MAPE	0.748266

Figure 6. The Forecast and its 95% Confidence Limits for the Training Data Set

According to Figure 6, the Theil Inequality Coefficient (U) was 0.006 with a minute bias proportion (0.0002). Since the U was closer to 0, it indicated that the predicting power of the model was good. According to the training data set MAPE is only 0.74%.

Validation of the Independent Data Set

Using best fitted model ARIMA (1,1,1) ARCH (2), the values from the 3rd of June to 14th of June was computed and compared with actuals



Forecast: EXCHANGE_RF	
Actual: EXCHANGE_RATE	
Forecast sample: 6/03/2024 6/14/2024	
Included observations: 10	
Root Mean Squared Error	7777.208
Mean Absolute Error	7083.726
Mean Abs. Percent Error	1.006750
Theil Inequality Coef.	0.005499
Bias Proportion	0.282862
Variance Proportion	0.657934
Covariance Proportion	0.059203
Theil U2 Coefficient	0.894292
Symmetric MAPE	1.002821

Figure 7. The Forecasts and its Confidence Limits for the Validation Data

According to the temporal variability of the observed series and the temporal variability of the forecasted series in Figure 7, the MAPE is only 1.002% and the Inequality Coefficient (U=0.064) is closer to 0 with a minute bias proportion (0.156), indicating that the model's predicting power was good.

Validation Data with Predicted Values and Percentage Errors

Table 3. Validation Data with Predicted Values and Percentage Errors

Date	Actual gold price (LKR)	Predicted gold price (LKR)	Percentage error
6/3/2024	703082.3	708575.9	-0.78
6/4/2024	708650.5	708681.3	0.00
6/5/2024	705045	708799.7	-0.53
6/6/2024	708608.4	708929.3	-0.05
6/7/2024	717719.3	709068.9	1.22
6/10/2024	695902.1	709217.3	-1.88
6/11/2024	697962.2	709373.4	-1.61
6/12/2024	702505.2	709536.2	-0.99
6/13/2024	703449.6	709705	-0.88
6/14/2024	700477.6	709879	-1.32

According to Table 2, the percentage errors of actual gold price and the predicated gold price of the validation data vary from -1.61% to 1.22%. These results indicate that the model can be validated for the independent data set.

Short-term Predictions

Table 4. Predicted Values for the Upcoming Days

Date	Gold price rate of ounce (LKR)
6/18/2024	709308.2
6/19/2024	709525.8
6/20/2024	709742.8

Using the best fitted model, the gold price of a troy ounce was predicted for the upcoming days. According to these results appropriate decisions can be made.

Conclusion & Recommendations

This research paper examined the temporal variability and forecasting accuracy of daily gold prices using ARIMA and ARCH models. The results confirmed that the ARIMA (111) ARCH (2) model is the best fit due to its significant coefficients, maximum log likelihood, and lowest AIC, SIC, and HQI values. The heteroskedasticity tests and correlograms confirmed that errors are identically and independently distributed, with no ARCH effect present.

Best fitted model,

$$Y_t = 212.34 + 0.879Y_{t-1} + e_t + 0.837e_{t-1}$$

- mean equation

$$\sigma_t^2 = 21054345 + 0.295\varepsilon_{t-1}^2 + 0.195\varepsilon_{t-2}^2$$

- Variance equation

The predictive power of the model was validated using an independent dataset, showing a minimal Mean Absolute Percentage Error (MAPE) of 1.002% and a Theil Inequality Coefficient (U) of 0.064. These results underscore the model's robustness in accurately predicting gold prices, with the validation data's percentage errors ranging between -1.61% and 1.22%.

Recommendations

Financial institutions and investors are encouraged to adopt the ARIMA (1,1,1) + ARCH (2) model for short-term forecasting of gold prices. The model's proven accuracy and low error rates make it a reliable tool for predicting market trends and making informed investment decisions.

When forecasting, it is important not to rely solely on past values. Subjective aspects must also be considered to improve the accuracy and relevance of the forecasts. As an example, the sudden rise in gold prices at the end of 2021 was propelled by increasing inflation fears worldwide, prompting investors to turn to gold as a hedge. Geopolitical tensions and a weakening U.S. dollar further bolstered demand for the precious metal, reinforcing its status as a safe-haven asset.

At the end of 2021, the gold price per ounce increased suddenly. To address this abrupt change, we can incorporate structural breaks into our time series model, which will enhance the accuracy of our predictions. By identifying and accounting for these structural breaks, the model can better capture the underlying dynamics of gold price movements, leading to more reliable forecasts.

By implementing these recommendations, stakeholders can leverage advanced time series modeling techniques to enhance their forecasting capabilities and make more strategic financial decisions.

Acknowledgement

I would like to express my deepest gratitude to the Head of the Department of Mathematics and Statistics and all the esteemed lecturers for their invaluable guidance and support throughout this project. Their expertise and encouragement have been instrumental in shaping my understanding and application of advanced statistical methods. This work would not have been possible without their dedication and commitment to excellence in education.

References

Athanasopoulos, R. J. (2021). *Forecasting: Principles and Practice.* Melbourne, Australia

Central Bank SriLanka. (2018). *Gold Price (in LKR)*. Retrieved from cbsl.gov: https://www.cbsl.gov.lk/en/rates-and-indicators/exchange-rates/daily-gold-rates

Hayes, A. (July 09, 2024). Gold vs. Stocks and Bonds. *Has Gold Been a Good Investment Over the Long Term*.

Khan, D. M. (2012). Forecasting of Gold Prices (Box Jenkins Approach). *International Journal of Emerging Technology and Advanced Engineering*, 3(3):662-670

Latifa Ghalayini, S. F. (April 2020). Modeling and Forecasting Gold Prices. pg 1-39.

Shalini Kakkar, D. P. (2023). *Forecasting of Gold Prices using ARIMA Model.* European Economic Letters.

Tassel, S. (1985). Forecasting the price of gold: A fundamentalist approach. *Journal of Atlantic Economic*, Volume 13:43-51.

Wee Chian Koh, J. B. (July 27, 2020). Gold price. *Gold shines bright throughout the COVID-19 crisis.*