Forecasting Daily Gold Price Using ARIMA Box-Jenkins Method

Fernando William Alexander

Study Program of Actuarial Science, Faculty of Business, President University, Bekasi, 17550, Indonesia
Corresponding author: fernando.alexander@student.president.ac.id

Abstract— In this research, the ARIMA time series model is utilized to make predictions of gold prices from March 1, 2025, to May 12, 2025. The prediction of the movement of the price of gold from May 13 to May 17, 2025, is intended. A number of ARIMA settings were tested manually whereby the order of differencing was designed to be two to deal with non-stationarity. Using AIC and BIC as criteria, the best models were the ARIMA (4,2,0) and ARIMA (4,2,1). These models produced short-term forecasts. The approach seeks transparency in model selection as well as interpretability in order to facilitate informed forecasting.

Keywords— Gold Price; Forecasting; Time Series; ARIMA; Model Selection

I. INTRODUCTION

Main objective of the analysis of a time series is to forecast its future value [1]. Forecasting is a tool as a decision-making that help businesses cope with the impact of the future's uncertainty by analyzing historical data [2]. There are many factors that affect the accuracy of the forecasting, including how much data is available, how similar the future is to its past, and sometimes our forecast affect the thing that we are trying to forecast [3]. Hence, many studies and tools are used for studying forecast or forecasting.

One of many methods of forecasting is using Autoregressive Integrated Moving Average (ARIMA). ARIMA model mainly used due to easy-to-use concept and utility of algorithms [4]. It is describing the autocorrelation of the data [3]. Main idea of the model is considering the data's dynamic relations over time, it is said that the future data dependent on past data [5]. Notation of ARIMA, which is ARIMA (p,d,q). The order p came from the Autoregressive (AR) model which explains that there is a regression of the variable against itself. While the q-order is from the Moving Average (MA) that explains about past forecast errors in a regression-like model. For the d-order, it came from the transformation of the data that was not stationary, how many times do the data need to be transformed in order to become stationary is the order of the d. Main focus of the stationary is to eliminating trend and seasonality [3].

Performing the ARIMA to testing is one of the ways to get know about how ARIMA model works. Applying the method to forecasting Gold Price daily price from 13th May 2025

– 17th May 2025 from the data since 1st March 2025 – 12th May 2025 with a total of 78 observations of the data practice and 5 observations for the data testing. This method is ideal for data with trends and seasonality, offering simplicity, clarity, and reliability [6]. In hope that, the accuracy of ARIMA testing can be shown and gain practical insight for future strategy.

II. METHODOLOGY

A. Box-Jenkins Method

Box-Jenkins method is used to forecasts time series data for a specific period of time [7]. It is suitable for the goal of forecasting the gold price for 5 days ahead. It has several steps, which identify, estimate, diagnose, refining, and forecasting. Or in other words, prepare the data, check if the data stationary, specify the model, estimating the parameters, analyzing the residuals, and forecasting the data. Data has to be stationary assumptions to be able to forecast [6]. Those data that are not stationary can still be done by using ARIMA (p,d,q) method.

1. Autoregressive (AR)

   AR model shows current data (Y_t) depending on data from previous period (Y_{t-p}) [6]. Equation of AR can be shown by:

   $$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi p Y_{t-p} + et \quad (1)$$

2. Moving Average (MA)

   MA is a process that we get when only a finite number of the ψ-weights are nonzero [8]. Expression MA can be shown as:

   $$Y_t = et - \theta e_{t-1} - \theta_2 e_{t-2} - \cdots - \theta_{t-q} e_{t-q} \quad (2)$$

3. Autoregressive Integrated Moving Average (ARIMA)

   As explained on the Introduction, ARIMA is a combination of AR and MA and using differencing $W_t = Y_t - Y_{t-1}$ [9].

   Equation of ARIMA can be shown by:

$$W_t = \phi_1 W_{t-1} + \phi_2 W_{t-2} + \cdots + \phi p W_{t-p} + et - \theta e_{t-1} - \theta_2 e_{t-2} - \cdots - \theta_{t-q} e_{t-q} \quad (3)$$

Wt = Differencing Process

Φp = AR Coefficient at order p

Θq = MA Coefficient at order q

e_t = Error at time t

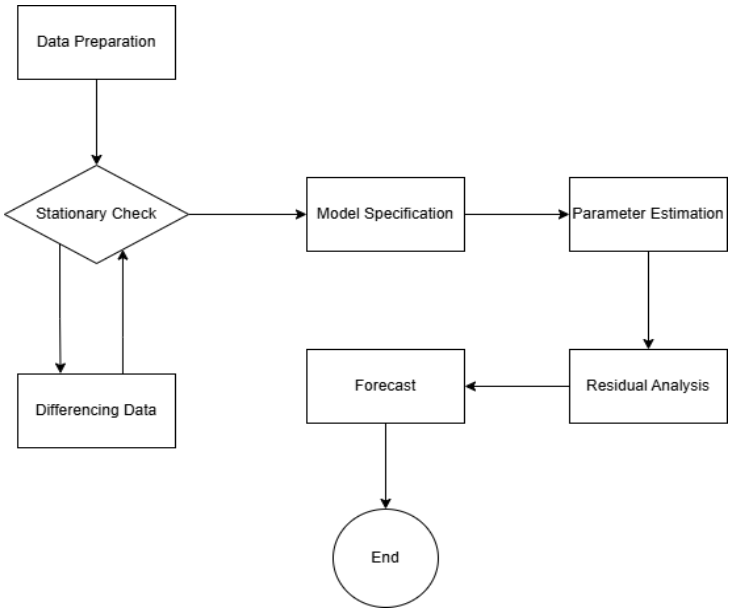

Figure 1. ARIMA Forecasting Steps

III. RESULTS AND DISCUSSION

A. Data Preparation

Data that are used for this analysis is Gold Price data obtained by data from OCBC from 3rd March 2025 until 17th May 2025 consists of 78 observations which are divided into 73 as data training and 5 data as data test. The list of data will be shown in Table 1, and the graph of the data will be shown in Figure 2.

Table 1.

Gold Price

Date	Price	Date	Price	Date	Price	Date	Price	Date	Price
3/1/2025	1574000	3/16/2025	1634000	3/31/2025	1697000	4/15/2025	1847000	4/30/2025	1900000
3/2/2025	1574000	3/17/2025	1636000	4/1/2025	1715000	4/16/2025	1875000	5/1/2025	1869000
3/3/2025	1581000	3/18/2025	1640000	4/2/2025	1712000	4/17/2025	1901000	5/2/2025	1851000
3/4/2025	1602000	3/19/2025	1652000	4/3/2025	1739000	4/18/2025	1905000	5/3/2025	1841000
3/5/2025	1607000	3/20/2025	1670000	4/4/2025	1711000	4/19/2025	1902000	5/4/2025	1841000

3/6/2025	1604000	3/21/2025	1668000	4/5/2025	1681000	4/20/2025	1902000	5/5/2025	1841000
3/7/2025	1593000	3/22/2025	1658000	4/6/2025	1681000	4/21/2025	1916000	5/6/2025	1863000
3/8/2025	1593000	3/23/2025	1658000	4/7/2025	1673000	4/22/2025	1971000	5/7/2025	1889000
3/9/2025	1593000	3/24/2025	1659000	4/8/2025	1676000	4/23/2025	1930000	5/8/2025	1895000
3/10/2025	1596000	3/25/2025	1653000	4/9/2025	1728000	4/24/2025	1907000	5/9/2025	1866000
3/11/2025	1585000	3/26/2025	1664000	4/10/2025	1761000	4/25/2025	1907000	5/10/2025	1866000
3/12/2025	1604000	3/27/2025	1671000	4/11/2025	1811000	4/26/2025	1900000	5/11/2025	1866000
3/13/2025	1614000	3/28/2025	1683000	4/12/2025	1842000	4/27/2025	1900000	5/12/2025	1846000
3/14/2025	1637000	3/29/2025	1697000	4/13/2025	1842000	4/28/2025	1897000		
3/15/2025	1634000	3/30/2025	1697000	4/14/2025	1847000	4/29/2025	1900000		

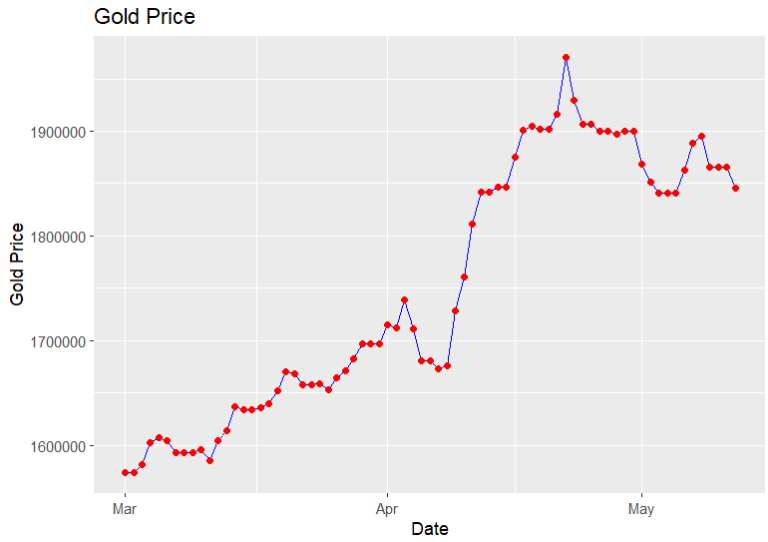

Figure 2. Plot of Gold Price

B. Stationary Check

Through Augmented Dickey-Fuller (ADF) Test method in RStudio with code adf.test(data). If the data has a p-value below 0,05, then the data considered stationary which is the requirement for the data to be processed. Value of p from first ADF Test is 0.8693, implies that it is not stationary. Therefore, differentiating the data is needed.

To make the data stationary, differencing is needed by diff(data) code in RStudio. It will be shown in Figure 3 the plot of the first differentiation between the data.

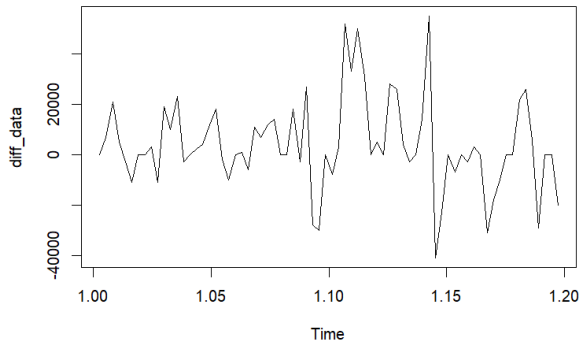

Figure 3. Plot of First Difference of Gold Price

After the first difference of the data, the p-value from ADF Test is 0.0547, which means it still cannot be assumed stationary. Differences one more time are needed, by diff(diff_data) we got the p-value below 0.01. Therefore, we can use the d-order for ARIMA model with value of 2

C. Model Specification

Determining the ARIMA model can be shown in the Auto-Correlation Function (ACF) and Partial Auto-Correlation Function (PACF) to show the q and p values, respectively. The plot of ACF and PACF from the data will be shown in Figure 4.

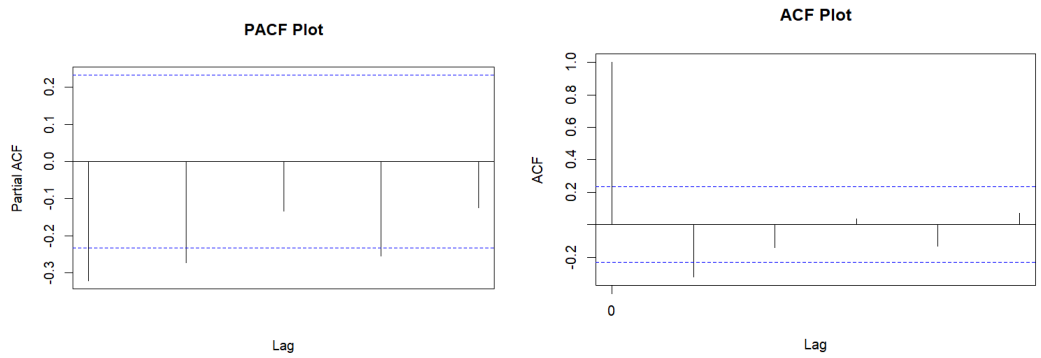

Figure 4. Plot of ACF and PACF Second Difference of Gold Price

Based on Figure 4, it is known that the cuts off at a lag time on ACF and PACF are 1 and 4, respectively. This implies that the order of p is determined as 4 and the order of q is 1 Arrangements of ARIMA model based on these parameters will be shown in Table 2.

Table 2 Model Specifications

No	ARIMA (p,d,q)	p	d	q
1	(0,2,0)	0	2	0
2	(1,2,0)	1	2	0
3	(2,2,0)	2	2	0
4	(2,2,0)	3	2	0
5	(4,2,0)	4	2	0
6	(0,2,1)	0	2	1
7	(1,2,1)	1	2	1
8	(2,2,1)	2	2	1
9	(3,2,1)	3	2	1
10	(4,2,1)	4	2	1

D. Parameter Estimation

ARIMA is Autoregressive Integrated Moving Average, where Autoregressive (AR) formula is denoted as φ with order p. On the other hand, Moving Average (MA) formula is denoted by θ with order q. Through summary(model), the value of each parameter can be obtained which will be shown in Table 3 and Log Likelihood and Akaike Information Criterion (AIC) will be shown in Table 4 can be obtained.

Table 3A

AR0 – AR4 Parameters

ARIMA (p,d,q)	AR(1)	AR(2)	AR(3)	AR(4)	LogLikelihood	AIC
(0,2,0)					-859.4	1720.807
(1,2,0)	-0.6791				-837.69	1679.386
(2,2,0)	-1.0961	-0.6044			-822.51	1651.023
(3,2,0)	-1.3272	-1.026	-0.3792		-817.26	1642.527
(4,2,0)	-1.4642	-1.3986	-0.8694	-0.3644	-812.48	1634.968

Table 3B

AR0-AR4-MA1 Parameters

ARIMA (p,d,q)	AR(1)	AR(2)	AR(3)	AR(4)	MA(1)	LogLikelihood	AIC

(0,2,1)					-1	-822.17	1648.35
(1,2,1)	-0.5558				-1	-809.34	1624.68
(2,2,1)	-0.8297	-0.4656			-1	-801.11	1610.22
(3,2,1)	-0.9447	-0.6722	-0.2448		-1	-799.06	1608.12
(4,2,1)	-1.0315	-0.8991	-0.5651	-0.3239	-1	-795.38	1602.75

E. Residual Analysis

According to [3], residual diagnostics are essential in time series modeling to ensure that the residuals resemble white noise, indicating a good model fit. Tests that are used to determine the best ARIMA model are Saphiro Test and the Ljung-Box Test. Data that has p-value in both tests is considered a 'model fit'. P-values for all models through these tests with code saphiro.test(model) and residual(model) will be shown in Table 5.

Table 5

Result of Residual Analysis

ARIMA (p,d,q)	**Ljung Box**	**Saphiro-Test**	**Result**
(0,2,0)	0.0000000000076	0.004919558	Not Passed
(1,2,0)	0.00000004137	0.027566061	Not Passed
(2,2,0)	0.006411	0.011067178	Not Passed
(3,2,0)	0.004772	0.001740822	Not Passed
(4,2,0)	0.008823	0.000806624	Not Passed

ARIMA (p,d,q)	**Ljung Box**	**Saphiro-Test**	**Result**
(0,2,1)	0.000000	0.002774907	Not Passed
(1,2,1)	0.000393	0.002080539	Not Passed
(2,2,1)	0.068750	0.000318991	Not Passed
(3,2,1)	0.011940	0.000582767	Not Passed
(4,2,1)	0.046420	0.005735577	Not Passed

As the information provided on Table 5, all the data are not considered as a model fit. Therefore, another method should be used to determine the best model. Using model's likelihood and a penalty term that increases with the number of parameters. Akaike

information Criterion (AIC) and Bayesian information Criterion (BIC) values evaluate how well the model fits the data while avoiding overfitting [3]. Therefore, choosing the best model for forecasting can be done by choosing the least value from AIC or BIC. Because lower the AIC or BIC provides the best trade-off between accuracy and overfitting. All AIC and BIC values will be shown on Table 6.

Table 6
AIC and BIC Value of ARIMA Model

ARIMA (p,d,q)	AIC	BIC	AIC Result	BIC Result
(0,2,0)	1720.807	1723.04	Failed	Failed
(1,2,0)	1679.386	1683.86	Failed	Failed
(2,2,0)	1651.023	1657.73	Failed	Failed
(3,2,0)	1642.527	1651.46	Failed	Failed
(4,2,0)	1634.968	1646.14	Best	Best

ARIMA (p,d,q)	AIC	BIC	AIC Result	BIC Result
(0,2,1)	1648.349	1652.82	Failed	Failed
(1,2,1)	1624.679	1631.38	Failed	Failed
(2,2,1)	1610.219	1619.16	Failed	Failed
(3,2,1)	1608.118	1619.29	Failed	Failed
(4,2,1)	1602.753	1616.16	Best	Best

From Table 6, the least value of BIC and AIC is on ARIMA (4,2,0) and ARIMA (4,2,1), respectively. The models with the lowest value will be forecasted, and the error for each model will be calculated.

F. Forecast

Table 7
Forecasting Value of ARIMA (4,2,0)

Date	Actual Value	Predicted Data	Lower Bound	Upper Bound
5/13/2025	1824000	1813932	1753431	1874433
5/14/2025	1826000	1812555	1743919	1881192
5/15/2025	1805000	1812919	1728340	1897498
5/16/2025	1828000	1805186	1694305	1916068
5/17/2025	1809000	1813376	1677712	1949040

Table 8
Forecasting Value of ARIMA (4,2,1)

Date	Actual Value	Predicted Value	Lower Bound	Upper Bound
5/13/2025	1824000	1804095	1758495	1849696
5/14/2025	1826000	1801865	1756263	1847466
5/15/2025	1805000	1810987	1764896	1857079
5/16/2025	1828000	1801230	1752579	1849881
5/17/2025	1809000	1799163	1749107	1849219

Forecasting values that are seen on Table 7 and Table 8 can be compared to determine the best ARIMA model by computing the Mean Square Error (MSE), Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE). MAPE is the average value of the absolute forecasting errors as a percentage of the respective data value. It measures relative performance hence it is the most useful measure compared to the accuracy of the forecast. Interpretation of MAPE such as <10% means highly accurate, 10-20% means good forecasting, 20-50% means reasonable forecasting, and >50% means inaccurate forecasting [10].

Table 9
Error Measured from ARIMA Models

ARIMA	(4,2,0)	(4,2,1)
MAE	11724.25	17326.86
RMSE	13299.96	19120.43

MAPE	0.64%	0.95%

Based on the values of each metric from Table 9. The ARIMA (4,2,0) has the least error by 0.64% from the actual data. Proving that ARIMA (4,2,0) is the most effective model for forecasting gold price. Referring to Table 3A, the model can be expressed in the form of equation:

$$W_t = -1.4642W_{t-1} - 1.3986W_{t-2} - 0.8694W_{t-3} - 0.3644W_{t-4} + e_t \quad (4)$$

By letting,

$$W_t = Y_t - Y_{t-1} \quad (5)$$

Then,

$$Y_t - Y_{t-1} = -0.14642(Y_{t-1} - Y_{t-2}) - 1.3986(Y_{t-2} - Y_{t-3}) - 0.8694(Y_{t-3} - Y_{t-4}) - 0.3644(Y_{t-4} - Y_{t-5}) + e_t$$

$$Yt = 0.85358Y_{t-1} - 1.25218Y_{t-2} + 0.5292Y_{t-3} + 0.505Y_{t-4} + 0.3644Y_{t-5} + e_t \quad (6)$$

Plot of ARIMA (4,2,0) will be shown in Figure 5 that visualize the stock price of Gold Price from 13th May 2025 – 17th May 2025

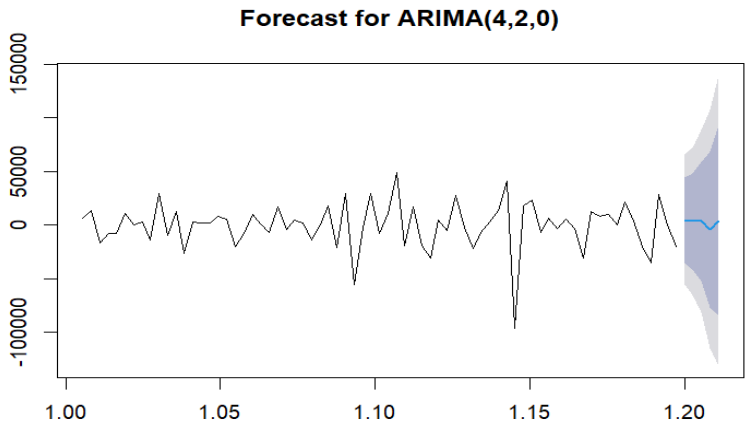

Figure 5 Plot of ARIMA (4,1,0)

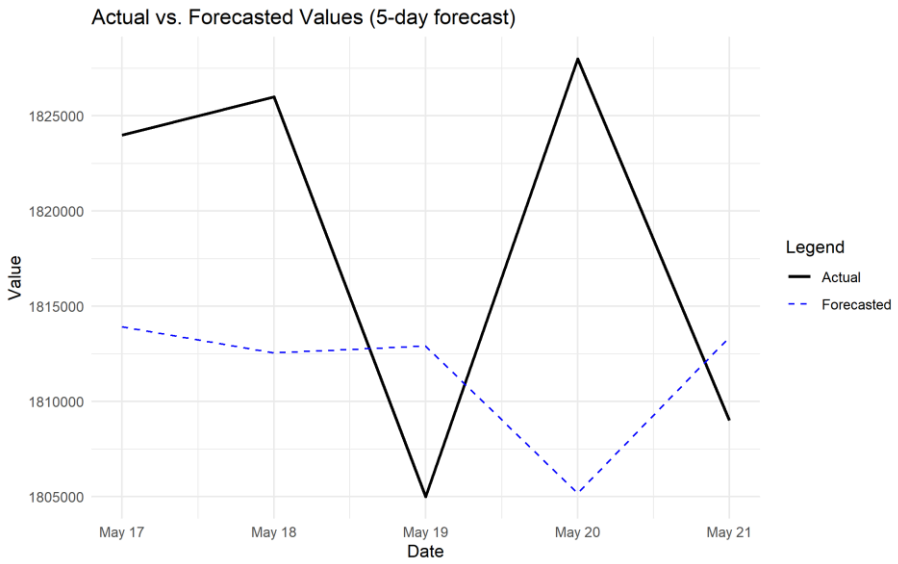

Figure 6. Comparison of the Actual and the Predicted Data

Conclusion

Through ARIMA test, by the best model. Gold price successfully predicted with a margin of error approximately 0.64%. Forecasting the data from 13th May 2025 – 17th May 2025 by data from 1st March 2025 to 12th May 2025. It was found that the best ARIMA Model is ARIMA (4,2,0) with the following equation:

$$Yt = 0.85358Y_{t-1} - 1.25218Y_{t-2} + 0.5292Y_{t-3} + 0.505Y_{t-4} + 0.3644Y_{t-5} + e_t$$

REFERENCES

[1] W. W. S. Wei, *Time Series Analysis Univariate and Multivariate Methods*. in 2nd Edition. Greg Tobin. [Online]. Available: https://civil.colorado.edu/~balajir/CVEN6833/lectures/wwts-book.pdf

[2] jeff Schmidt, "Forecasting," Corporate Finance Institute. Accessed: May 04, 2025. [Online]. Available: https://corporatefinanceinstitute.com/resources/valuation/forecasting/

[3] R. J. Hyndman and G. Athanasopoulus, *Forecasting: Principles and Practice (3rd ed)*. Accessed: May 04, 2025. [Online]. Available: https://otexts.com/fpp3/

[4] W. Hariadi and Sulantari, "Application of ARIMA Model for Forecasting Additional Positive Cases of Covid-19 in Jember Regency," vol. 1, no. 1, Apr. 2021.

[5] I. Svetunkov, *Chapter 8 Introduction to ARIMA | Forecasting and Analytics with the Augmented Dynamic Adaptive Model (ADAM)*. Accessed: May 05, 2025. [Online]. Available: https://openforecast.org/adam/ARIMA.html

[6] A. I. S. Wardhani and M. R. Yudhanegara, "Forecasting Weekly Stock Price of PT. Aneka Tambang Tbk (ANTM) Using ARIMA Box-Jenkins Method," *JAFRM*, vol. 3, no. 2, Jan. 2025.

[7] "Box-Jenkins Methodology for ARIMA Models," GeeksforGeeks. Accessed: May 05, 2025. [Online]. Available: https://www.geeksforgeeks.org/box-jenkins-methodology-for-arima-models/

[8] J. Saputra, "Forecasting Stock Price of PT. GoTo Gojek Tokopedia Tbk (GOTO.JK) using ARIMA Box-Jenkins Method," *Indones. J. Appl. Math. Stat.*, vol. 1, no. 2, Art. no. 2, Dec. 2024, doi: 10.71385/idjams.v1i2.14.

[9] J. D. Cryer and K.-S. Chan, "Time-Series-Analysis/time-series-analysis-with-applications-in-r-cryer-and-chan.pdf at master · svbergmann/Time-Series-Analysis," GitHub. Accessed: May 05, 2025. [Online]. Available: https://github.com/svbergmann/Time-Series-Analysis/blob/master/time-series-analysis-with-applications-in-r-cryer-and-chan.pdf

[10] C. D. (Colin D. Lewis, *Industrial and business forecasting methods : a practical guide to exponential smoothing and curve fitting*. London ; Boston : Butterworth Scientific, 1982. Accessed: May 05, 2025. [Online]. Available: http://archive.org/details/industrialbusine0000l