Ailyn Song
Forecasting Time Series
Professor Clifford M. Hurvich
Nov. 6, 2024

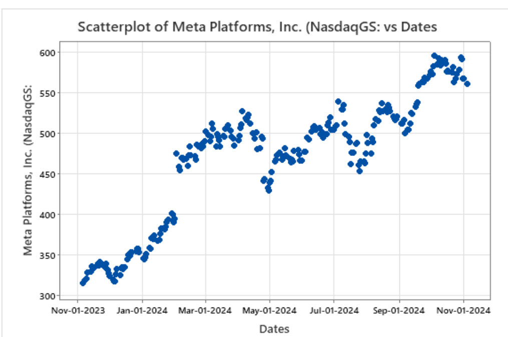
Project 1: Forecasting the Meta Stock Using ARIMA

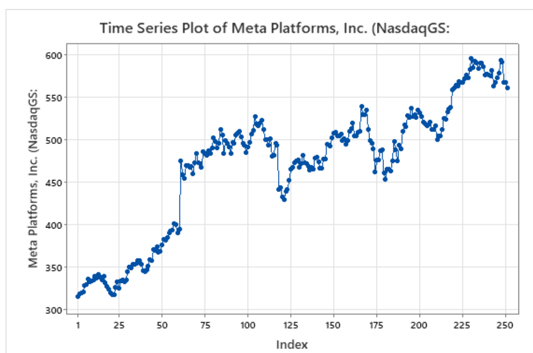
Introduction & Data Description

The dataset used in this project consists of the stock prices of Meta (formerly Facebook, Inc.), a leading technology company, over the past year. This project aims to analyze and forecast Meta's stock price using ARIMA modeling, based on historical data.
The dataset includes daily observations, totaling 251 data points, covering the period from Nov-06-2023 to Nov-04-2024. The data was sourced from [Capital IQ](). Given the short time span, no seasonal adjustment was applied, as daily stock prices typically do not exhibit strong seasonal patterns in such a narrow time window.

Plotting the Data

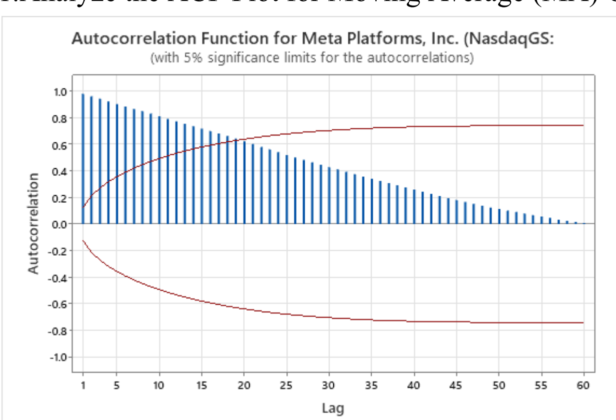
I started by plotting the Meta stock price data from the past year. The plot displayed a general upward trend, with fluctuations that increased as the stock price level rose. This pattern does not show exponential trend, nor does it is suggests that the data may be too level-dependent with high variability. After the log transformation, the scatter plot and time series do not change much either. So I decided not to log the data.





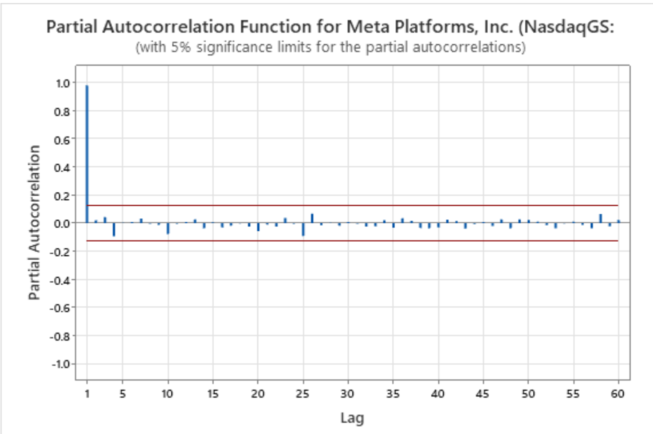
Model Selection

1.Analyze the ACF Plot for Moving Average (MA) Order (q)



In the ACF plot provided, we observe a gradual decay or hang over the lags rather than a sharp cutoff. This is indicative of an autoregressive (AR) process because the autocorrelations gradually decrease as the lags increase. Since there's no immediate cutoff in the ACF, it does not indicate an MA process, meaning q could initially be set to 0.

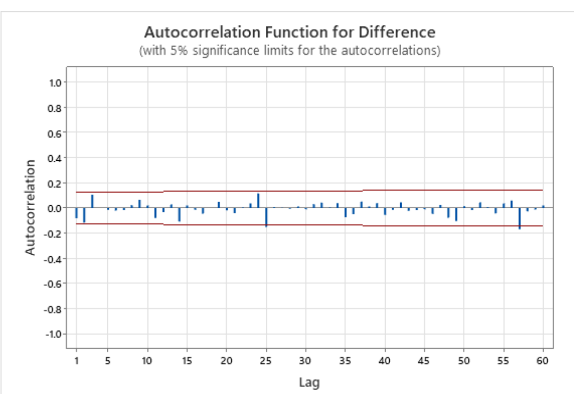
2.Analyze the PACF Plot for Autoregressive (AR) Order (p):

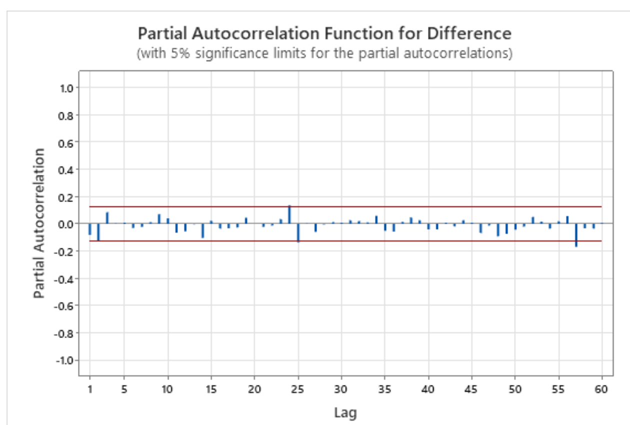


The PACF plot also has a significant spike at lag 1, with a few smaller significant values at later lags and most within confidence interval. This behavior is typically associated with a low-order AR component. Since the PACF shows a prominent spike at lag 1 and tapers off, I might consider an AR(1) or AR(2) model, meaning p=1 or p=2.

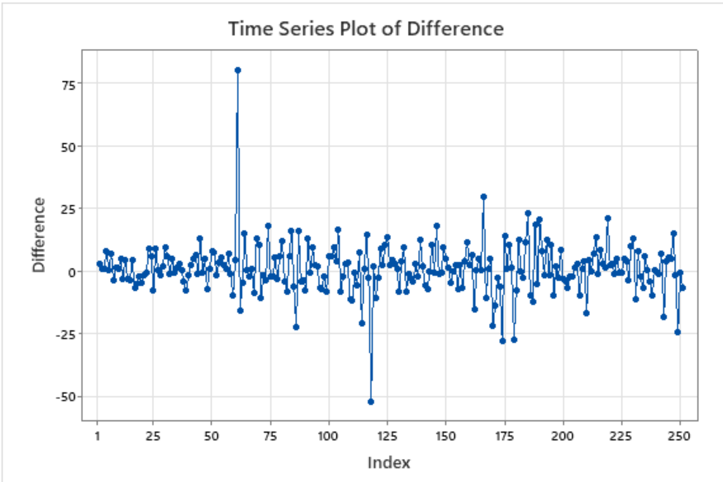
3. Differencing Order (d)

The ACF plot exhibits a gradual decay or slow "hanging" pattern rather than a sharp drop-off. This often indicates that the time series is non-stationary, as it implies a persistence in the autocorrelation structure across multiple lags. In the PACF plot, the spike at lag 1 is close to 1, indicating a high degree of autocorrelation at this lag.





After differencing once, the ACF for difference shows values that generally hover around zero, with minor fluctuations but no significant spike after lag 1. This behavior indicates that the first differencing has helped in achieving stationarity. The PACF follows this pattern, which is consistent with a stationary process after differencing. Since both ACF and PACF are mostly within significance bounds after the first differencing, the series does not appear to be over-differenced, which would show an alternating pattern or overly large negative correlations at early lags, which is not evident here.

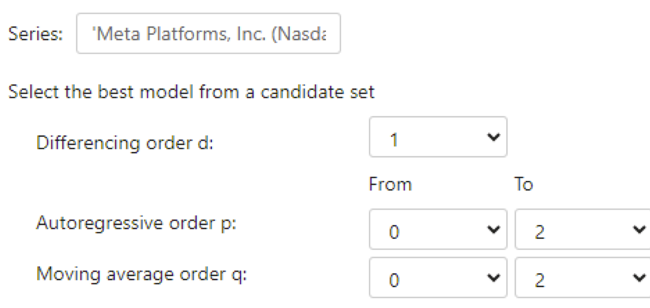


The time series plot of the differenced data confirms this, as it exhibits random fluctuations around zero with no visible patterns, aside from a few noticeable spikes. This randomness suggests that first-order differencing successfully transformed the series into a stationary one.

These indicators together point towards non-stationarity in the original series. Therefore, setting d=1 and applying a first difference transformation will likely make the series stationary for ARIMA.

Use AICc Values to Compare Models

In Forecast with Best ARIMA Model, we first test the one with constant:



Model Selection

Model (d = 1)	LogLikelihood	AICc	AIC	BIC
p = 2, q = 0*	-936.245	1880.65	1880.49	1894.58
p = 0, q = 2	-936.667	1881.50	1881.33	1895.42
p = 1, q = 2	-935.738	1881.72	1881.48	1899.08
p = 0, q = 1	-937.913	1881.92	1881.83	1892.39
p = 0, q = 0	-938.977	1882.00	1881.95	1889.00
p = 2, q = 1	-937.018	1882.20	1882.04	1896.12
p = 1, q = 0	-938.147	1882.39	1882.29	1892.86
p = 2, q = 2	-935.429	1883.20	1882.86	1903.99
p = 1, q = 1	-937.536	1883.24	1883.07	1897.16

** Best model with minimum AICc. Output for the best model follows.*

Then test the same setting with no constant:

Model Selection

Model (d = 1)	LogLikelihood	AICc	AIC	BIC
p = 2, q = 0*	-937.895	1881.89	1881.79	1892.35
p = 0, q = 0	-940.091	1882.20	1882.18	1885.70
p = 2, q = 1	-937.018	1882.20	1882.04	1896.12
p = 0, q = 1	-939.288	1882.62	1882.58	1889.62
p = 0, q = 2	-938.294	1882.69	1882.59	1893.15
p = 1, q = 2	-937.284	1882.73	1882.57	1896.65
p = 1, q = 0	-939.447	1882.94	1882.89	1889.94
p = 2, q = 2	-936.898	1884.04	1883.80	1901.40
p = 1, q = 1	-939.040	1884.18	1884.08	1894.64

** Best model with minimum AICc. Output for the best model follows.*

The table summarizes the results for various (p, q) configurations. The model with p=2 and q=0 (with a constant) provided the lowest AICc, indicating it as the best-fitting model among those tested.

Model Estimation and Equation

The final ARIMA(2, 1, 0) model can be represented by the equation:

Let Wt be the differenced series: $X_t - X_{t-1}$

$$W_t = \text{Constant} + \phi_1 W_{t-1} + \phi_2 W_{t-2} + \epsilon_t$$

Here, ϕ1and ϕ2 are autoregressive coefficients estimated by the model, which are listed in the "Final Estimates of Parameters" section.

Final Estimates of Parameters

Type	Coef	SE Coef	T-Value	P-Value
AR 1	-0.0919	0.0632	-1.45	0.147
AR 2	-0.1237	0.0632	-1.96	0.052
Constant	1.196	0.651	1.84	0.067

Differencing: 1 Regular
Number of observations after differencing: 250

The chosen model, ARIMA(2, 1, 0), was estimated, yielding the following equation:

$$X_t - X_{t-1} = 1.196 - 0.0919(X_{t-1} - X_{t-2}) - 0.1237(X_{t-2} - X_{t-3}) + \epsilon_t$$

Model Diagnostics

1.Ljung-Box Test:

Modified Box-Pierce (Ljung-Box) Chi-Square Statistic

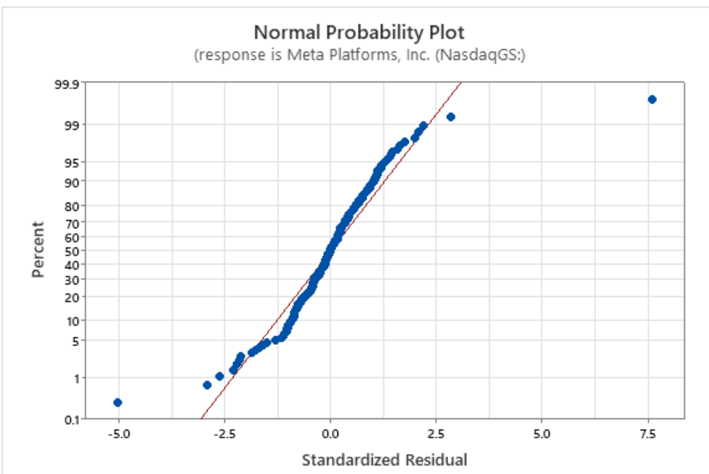
Lag	12	24	36	48
Chi-Square	5.36	13.96	23.13	29.53
DF	9	21	33	45
P-Value	0.802	0.871	0.900	0.964

The Chi-Square statistic increases with the lag, as expected, because it accumulates the effect of each successive lag's residual correlation.

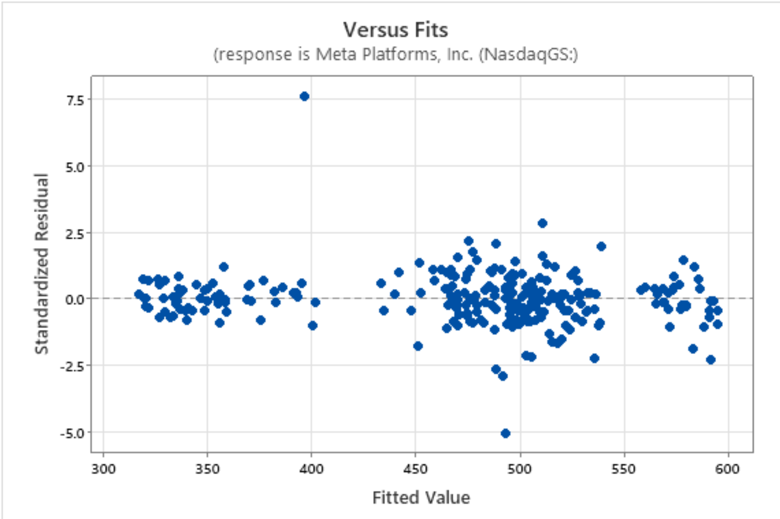
The degrees of freedom also increase with the lag (9, 21, 33, and 45). This is typical in the Ljung-Box test since DF corresponds to the lag number minus any parameters in the model.

Since all p-values are greater than 0.05, we fail to reject the null hypothesis of the Ljung-Box test at each lag level. This suggests that there is no significant autocorrelation in the residuals up to lag 48. The model appears to be adequate.

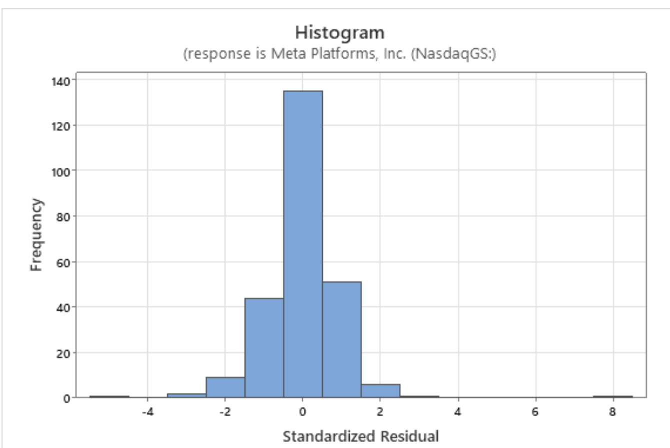
Residual Analysis



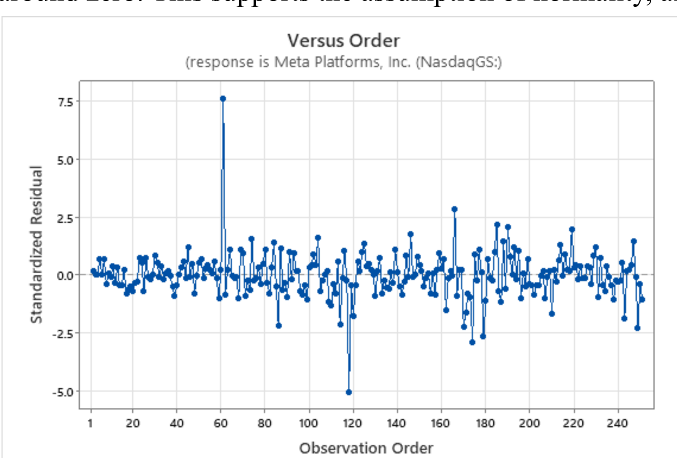
The Normal Probability Plot shows how well the residuals follow a normal distribution. Most of the points lie close to the line, indicating approximate normality. However, there are some outliers at the extremes, which could suggest occasional deviations from normality.



The scatter plot of residuals versus fitted values appears to be roughly centered around zero, with no obvious pattern. This indicates that the residuals are randomly distributed, which suggests that the model does not suffer from non-linearity or heteroscedasticity.

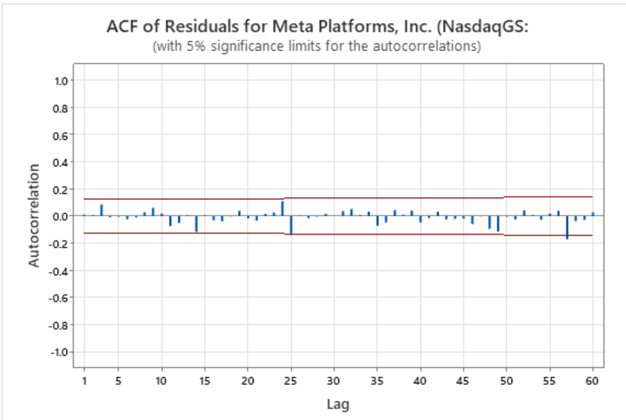


The histogram shows the distribution of residuals, which appears to be fairly symmetric and centered around zero. This supports the assumption of normality, although there may be slight deviations.



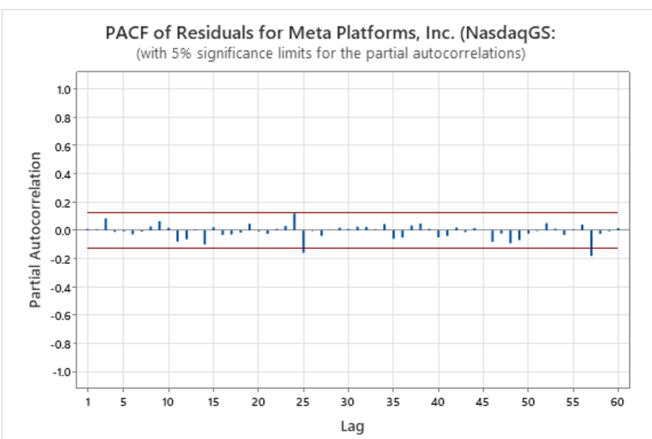
The time sequence plot of the residuals does not show any apparent trends or periodic patterns, and the residuals oscillate around zero. This suggests that the residuals are uncorrelated over time, which is ideal for a well-fitted time series model.

ACF (Autocorrelation Function) of Residuals:



The ACF plot shows most of the residual autocorrelations are within the 95% confidence interval limits, indicating that there is no significant autocorrelation left in the residuals. This is a positive sign, suggesting that the model has adequately captured the data's patterns.

PACF (Partial Autocorrelation Function) of Residuals:



The PACF plot shows, similarly, most values lie within the confidence interval limits. This further supports the absence of significant autocorrelation.

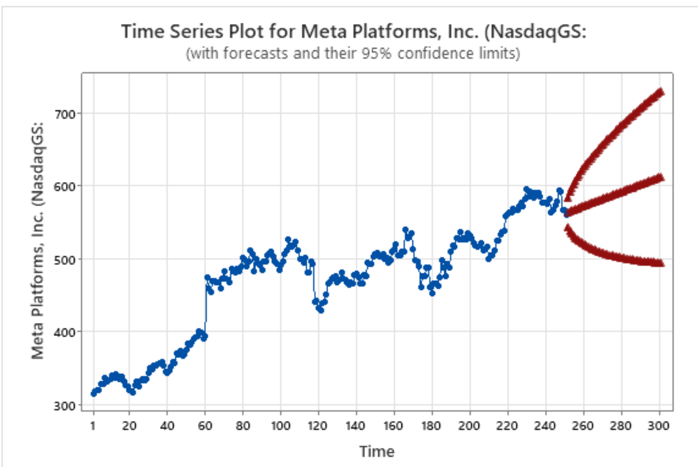
Summary of Diagnostic Checking:

- The residual diagnostics suggest that the model is appropriate for the data. The residuals are approximately normally distributed (as indicated by the Normal Probability Plot and Histogram) and do not show any strong patterns (as shown in the Residuals vs. Fitted plot and the Residuals vs. Order plot).
- Both the ACF and PACF of the residuals indicate no significant autocorrelations, confirming that the model has accounted for the autocorrelation in the data.

- Potential Issues: There are some outliers observed in the Normal Probability Plot and potentially a few high residuals in the time sequence plot. These may indicate occasional deviations from the model's assumptions. However, since these outliers are few, they may not significantly affect the model's overall performance.

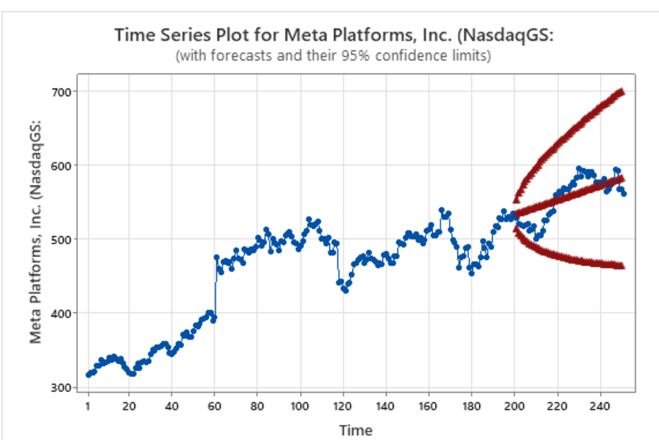
Forecasting

Using the ARIMA(2,1,0) model, I generated forecasts for the next 50 periods with 95% confidence intervals.



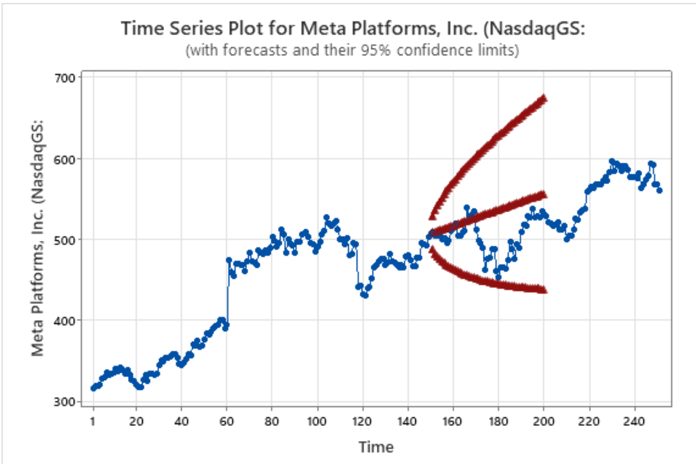
The forecasted values follow the general trend of the data series, showing a reasonable continuation of the existing data. This suggests that the model has captured the primary movement patterns in the data.

Backtesting with a lead time of 50 and an origin of 200:



The forecasted values appear to follow the existing trend in the data fairly well. The model projects the general upward movement observed near the origin (around time 200), indicating that it has captured the underlying pattern of the data. The 95% forecast interval is not too wide or narrow as it captures all the data in the range. Although the intervals do widen with time, they do not appear excessively wide given the observed variability in the original data. The width seems to appropriately balance caution with the inherent uncertainty of future values.

Backtesting with a lead time=50, with a origin=150:



The forecast from origin 150 shows good alignment with the actual values in the initial steps, following the trend fairly well. Similarly, the interval is not too wide or narrow. The data is captured by the interval and sometimes approaches the edge.

Both backtesting results show strong short-term performance and reasonable confidence intervals that widen as the forecast horizon extends.

Conclusion

Overall, this ARIMA model is suitable for short- to medium-term forecasting of Meta stock, where it can provide reasonably accurate predictions and realistic confidence intervals. For longer-term forecasts, however, the model may benefit from additional factors or adjustments to improve its stability and reliability, especially in capturing shifts in stock patterns over extended periods.