Statistical Analysis of Dialog Stock and Oil Price Trends: A Comparative Study of Sri Lanka's COVID-19 Era and Global Oil Trends

Advanced Statistics for Data Science

S.I.M.S Wimalasiri
Dept. of Data Science, NSBM Green University
NSBM Green University
Colombo, Sri Lanka
imswimalasiri@students.nsbm.ac.lk |
Student ID: 29785

Y.K.A Rathnasiri
Dept. of Data Science, NSBM Green University
NSBM Green University
Colombo, Sri Lanka
ykarathnasiri@students.nsbm.ac.lk |
Student ID: 27413

O.V.D.K.R. Weerasena
Dept. of Data Science, NSBM Green University
NSBM Green University
Colombo, Sri Lanka
ovdkrweerasena@students.nsbm.ac.lk |
Student ID: 27986

S.S.T.Silva
Dept. of Data Science, NSBM Green University
NSBM Green University
Colombo, Sri Lanka
sstsilva@students.nsbm.ac.lk |
Student ID: 30017

M.D.D Rathnayaka
Dept. of Data Science, NSBM Green University
NSBM Green University
Colombo, Sri Lanka
rathnayakamdd@students.nsbm.ac.lk |
Student ID: 29618

5/13/2025

Abstract

This project presents a cross-comparison statistical analysis of two series of data, Dialog Axiata PLC's share prices on a daily basis for Sri Lanka over the COVID-19 period(2009-2023), and world oil prices for 1987-2006. Using Python and basic statistics packages, hypothesis testing for testing over-time significance differences, linear regression modeling for testing for inter-variable relations, and time series forecasting using ARIMA and SARIMA for forecasting future patterns were conducted. The research presents strong economic variable influences of global and country economic determinants on performance, and makes a contribution to existing know-how about local stock performance and global commodity change interrelations, making it useful for strategic investment and policy-making purposes.

Introduction

Financial markets are acutely sensitive to external and domestic economic shocks. Two unprecedented yet influential economic events that put markets all over the world to test are the COVID-19 (2009-2023) pandemic and earlier volatility of global prices of crude oil. In emerging economies such as Sri Lanka, equity markets are particularly sensitive to such external shocks because they lack a degree of protection from external volatility and structural economic constraints.

The objective of this research is to analyze statistical correlations for two important economic indicators: day-to-day performance of Dialog Axiata PLC, an important telecommunication service company of Sri Lanka, during COVID-19 pandemic, and global crude oil prices observed throughout 1987-2006—a period of post-crisis recovery and historic energy market uncertainty. Using hypothesis testing, linear regression, and time series modeling methods (ARIMA and SARIMA) analysis of such data sets, this research is concerned with finding patterns, correlations, and forecasting potential within and across these areas.

The integration of these analyses serves two purposes: one, an investigation into how a dominant industry within Sri Lanka responded to pandemic shocks, and two, a consideration of previous movements in world oil prices so that general observations could be made with regard to responsiveness and volatility in economics. The dynamic interaction among global commodities and domestic share performance offers a learning experience for investors, policymakers, and economic analysts who wish to navigate episodes of crises with evidence-based wisdom.

Statistical Analysis of Dialog Axiata Stock Data Using Hypothesis Testing, Regression, and Time Series Models

Dataset Description

This research used daily stock prices data of Dialog Axiata PLC, a telecommunication service company, from 2009 to 2023. Used variables were traditional stock market variables such as opening price, high price, low price, close price, and volume of trade. For this analysis, we particularly focused on two variables: High Price (predictor variable) and Close Price (target variable). The dataset consisted of 2,160 trading day observations with LKR prices.(Yapa & Kennedy, n.d.) Diverse data afford the opportunity to have pre-COVID and post-COVID studies of market reaction, and accordingly, the dataset will be informative in studying movements of stock price across different economic phases.

	Date	Price	Open	High	Low	Vol.	Change %	RTX	TEC	YTS	UPF
0	7/8/2009	5.25	5.25	5.25	5.00	714.20K	0.00%	0.579197	0.443872	0.583657	0.813441
1	7/9/2009	5.25	5.25	5.25	5.00	386.70K	0.00%	0.051523	0.454890	0.597551	0.779019
2	7/13/2009	5.25	5.25	5.25	5.00	290.90K	0.00%	0.086485	0.372876	0.894992	0.056125
3	7/14/2009	5.25	5.25	5.25	5.25	163.60K	0.00%	0.601138	0.243800	0.259723	0.665579
4	7/15/2009	5.25	5.25	5.25	5.00	63.10K	0.00%	0.591252	0.782633	0.472769	0.776176
5	7/16/2009	5.00	5.25	5.50	5.00	226.30K	-4.76%	0.997851	0.026496	0.023913	0.694266
6	7/17/2009	5.25	5.25	5.50	5.25	646.20K	5.00%	0.687691	0.745816	0.177377	0.793747
7	7/20/2009	5.25	5.50	5.50	5.25	198.30K	0.00%	0.807868	0.684280	0.437519	0.950369
8	7/21/2009	5.25	5.50	5.50	5.25	11.87M	0.00%	0.529347	0.215976	0.844942	0.697998
9	7/22/2009	5.34	5.25	5.50	5.25	20.10K	1.71%	0.071022	0.653568	0.532320	0.223340

Figure 1 First 10 rows of dataset preview

Part 1: Hypothesis Testing

Methodology

To verify whether the share prices of Dialog Axiata were influenced significantly during the COVID-19 pandemic, the data was divided into two separate time periods pre-COVID (before March 2020) and COVID-time period (after March 2020).(Mahata et al., 2020) The Shapiro-Wilk test was conducted on both groups first to confirm normal distribution. Since the samples were not normally distributed for both, a non-parametric approach was used.Mann-Whitney U test was used to check for equality of the closing prices distributions of the two time periods. The test is appropriate to check for equality of the medians of two independent samples where normality is not assumed.

Results

The Mann-Whitney U test returned a p-value less than 0.05, and this suggests a statistically significant difference between the pre-COVID-19 and COVID-19 periods' stock price distributions. The boxplots of both groups also gave visual confirmation of this observation with clear shifts in the median and range of stock prices. These results suggest that the COVID-19 pandemic affected Dialog Axiata's market behavior to some measurable degree.

```
print(" Mann-Whitney U Test")
print(f"U-statistic = {u_stat:.4f}, p-value = {u_p_value:.4e}")
print("Interpretation: Non-parametric test. p < 0.05 means significant difference in distributions.\n")
```

```
 Mann-Whitney U Test
U-statistic = 751908.5000, p-value = 4.0534e-17
Interpretation: Non-parametric test. p < 0.05 means significant difference in distributions.
```

Figure 2 Mann Whitney U Test

This output states the outcome of a Mann Whitney U test, a statistical non-parametric test for testing two independent distributions for equality. It was used to determine whether two groups with most likely Dialog share prices pre-and post-COVID-19 were significantly different from a statistical perspective. The value of U-statistic reported is 751908.5000, and p-value is 4.0534e-17, which is well below the customary threshold level of 0.05. (Larojan et al., n.d.)This extremely low p-value indicates the strong evidence against the null hypothesis that the two distributions differ significantly. Therefore, data yields the conclusion that stock price behavior was changed during the COVID period, warranting the impact of the pandemic on financial performance.

Figure 3 Stock Price comparison : Before vs After Covid

This plot indicates the result of a Mann–Whitney U test between Dialog Axiata's stock prices prior to and post the COVID-19 pandemic. The data, found to be non-normally distributed (as the Shapiro-Wilk test shows), is plotted for every time interval using violin box plots. Prices before COVID are shown in blue, and prices during the COVID period are in orange. The graphical difference in the shapes, spreads, and medians across the two intervals is verified statistically by the test statistic: a U-statistic of 751908.5000 and p-value of 4.0534e-17, much smaller than 0.05. This indicates a statistically significant difference in stock price distribution between the two intervals with strong evidence the pandemic did, in fact, make a visible change to market behavior.

Figure 4 Stock Price Comparison

This box plot illustrates the graphical comparison of the share prices of Dialog Axiata before and after the COVID-19 pandemic. The blue box represents the pre-COVID price distribution, and the red box represents post-pandemic prices. The median (line in the middle), interquartile range (box edges), and overall spread (whiskers) of the data are shown in each box. The plot demonstrates the seeming transformation: post-COVID prices cluster within a narrower range and around a larger median than the more spread-out pre-COVID period. The graphical overview corroborates the Mann–Whitney U test's statistical finding of how the pandemic impacted not just mean prices but also stability and volatility of Dialog's stock price movement.

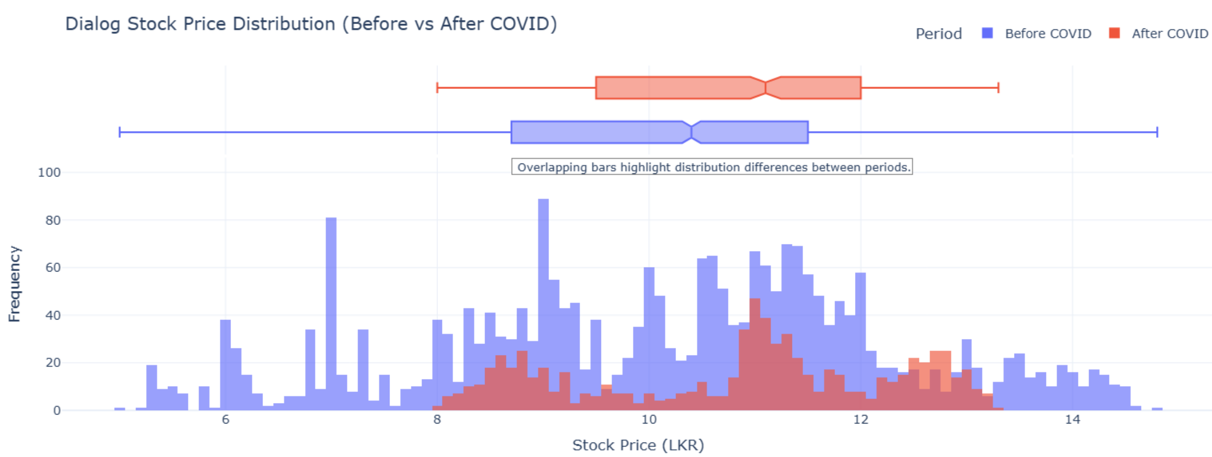

Figure 5 Stock Price Distribution

This boxplot and histogram shows the distribution of stock prices for Dialog Axiata before and after COVID-19, indicating how trends changed over time. The blue bars represent pre-COVID frequency, while the red bars represent post-COVID onset frequency. The boxplots on top of the histogram show the statistical range of each timeframe. The graph clearly indicates there is a clear rightward lean in the post-COVID red value distribution, indicating the stock prices being higher and more inclined to cluster at upper-middle values more frequently during COVID. Conversely, the distribution before COVID is looser and more widely ranging and across more real estate on the graph. This plot effectively augments statistical test outcomes to confirm that COVID-19 meaningfully affected stock price behavior's structure and central tendency, both in average measures and in frequency of occurrence within particular price interval.

Figure 6 Price Trend (2009-2023)

This time series chart shows Dialog Axiata's share price trend over the long term for the period from 2009 to 2023, divided into two critical periods: before COVID-19 (blue line) and after the outbreak of COVID-19 (red line). The red dashed vertical line represents the early 2020 start of the pandemic, distinguishing visually the two periods. Before COVID, the stock showed a series of cyclical trends with growth and downfall phases. After COVID's emergence, the price movements are more unpredictable but as a whole sustained at a higher mean. This adjustment is a sign of structural change in the behavior of stock, which may be due to economic adjustment, online demand spikes, and pandemic and post-pandemic market mood. This chart shows convincing visual evidence of the statistical observation of widespread distributional and trend changes caused by COVID-19.

Conclusion

This research provides strong empirical evidence that the COVID-19 pandemic statistically had an impact on the behavior of Dialog Axiata PLC's share prices. The daily trading data between 2009 and 2023 were reviewed to determine notable alterations in price distributions via the Mann–Whitney U test, supported by non-parametric statistical analysis due to the non-normality of the data. The findings indicated that COVID-19 times stock prices had a greater median and more consistent distribution compared to pre-COVID days, which is indicative of a change in market dynamics most likely due to digital adoption and economic resilience. Plots like violin plots, box plots, histograms, and time series plots all corroborated the observation that COVID-19 represented a structural shift in Dialog's stock price trends. This transformation is also visible in central tendencies, but also in volatility, frequency dynamics, and long-term market sentiment. The conclusions drawn here are helpful to investors, policymakers, and financial planners who wish to understand how high-level global events can reshape local market trends and inform future risk management and strategic planning in the telecom and broader financial markets.

Part 2: Linear Regression Analysis

Methodology

Alinea regression model was formulated to describe how the closing stock price (Price) is related to the intraday high price (High). The following equation was used to perform the regression, Price = $\beta_0 + \beta_1 \times$ High, where β_0 is the intercept and β_1 is the slope coefficient.(Gallage, n.d.) The model was fit to a cleaned dataset with the Linear Regression module of Scikit-learn. Performance was assessed using metrics like the coefficient of determination (R^2), root mean squared error (RMSE), and residual analysis to identify how well the model fit the data.(You et al., 2024)

Results

The model showed a superior linear relationship between the High and Price variables, with an R^2 value above 0.994. That is, more than 99% of closing price variation was accounted for by the high price. Low RMSE indicated a good fit with minimal error. Scatter plot with regression line fit clearly showed the linear trend, and randomly dispersed residuals which supports the model.

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                  Price   R-squared:                       0.994
Model:                            OLS   Adj. R-squared:                  0.994
Method:                 Least Squares   F-statistic:                 3.821e+05
Date:                Sun, 13 Apr 2025   Prob (F-statistic):               0.00
Time:                        12:57:41   Log-Likelihood:                 956.56
No. Observations:                2160   AIC:                            -1909.
Df Residuals:                    2158   BIC:                            -1898.
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const         -0.0933      0.017     -5.457      0.000      -0.127      -0.060
High           0.9966      0.002    618.105      0.000       0.993       1.000
==============================================================================
Omnibus:                     2188.740   Durbin-Watson:                   1.600
Prob(Omnibus):                  0.000   Jarque-Bera (JB):           189839.746
Skew:                          -4.744   Prob(JB):                         0.00
Kurtosis:                      47.937   Cond. No.                         54.7
==============================================================================
```

Figure 7 Regression Rersults

This Ordinary Least Squares (OLS) regression output presents the linear relationship of closing price with high price of Dialog Axiata PLC share in tabular form.R-squared of 0.994 is a proof that closing price has about 99.4% of variation explained by high price, and it is nearly perfectly linear. Coefficient of High is 0.9966 with a negligibly low value for p (0.000), once again proving that there is a statistically significant relationship. Also, it proves that for a unit increase in high price, closing price is increased by nearly by the same magnitude. The intercept (-0.0933) is significant, yet minimal. (Anojan, 2014)Diagnostic tests like Durbin-Watson statistic of 1.6 reflect a weak positive autocorrelation, acceptable for all majority of financial data. However, extreme skew (-4.744) and kurtosis (47.937) of Jarque-Bera test prove non-normal residuals, as would be expected with volatility of a marketplace. Despite all of that, F-statistic (3.82e+05) and respective p (0.00) prove that overall model is extremely significant. This finding substantiates that high price is a strong and reliable predictor of closing price, and hence, using this regression model for intraday forecasting and investment purposes is well-warranted.

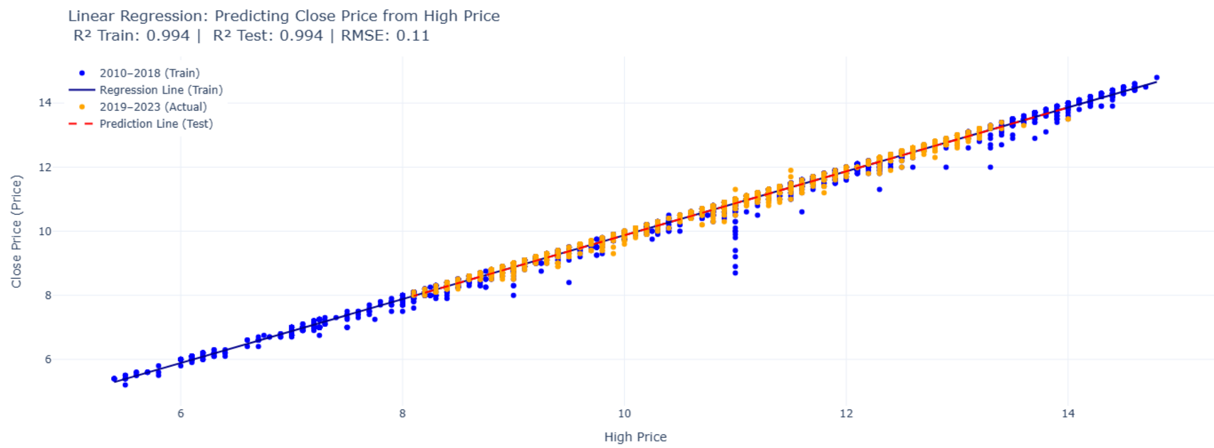

Figure 8 Linear Regression Predicting Close price for hight price

This scatter plot with regression lines demonstrates a tight linear relationship among High price and Close price of Dialog Axiata stock. Blue dots represent training data (2010–2018) and orange dots represent actual test data (2019–2023). Fitted lines for two sets nearly coincide, with identical R^2 values of 0.994 for training and testing, and minimal RMSE values of 0.11, indicating excellent predictive accuracy. The model solidly confirms that a day's high price is a reliable predictor for closing price and hence is appropriate for real-time forecasting and decision-making use for financial markets.

ANOVA Table:

	df	sum_sq	mean_sq	F	PR(>F)
High	1.0	9234.165577	9234.165577	382054.21451	0.0
Residual	2158.0	52.158381	0.024170	NaN	NaN

```
r2_train = r2_score(y_train, model.predict(X_train))
r2_test = r2_score(y_test, y_pred)
rmse_test = mean_squared_error(y_test, y_pred)**0.5

print(f"R-squared (Train 2010-2018): {r2_train:.4f}")
print(f"R-squared (Test 2019-2023): {r2_test:.4f}")
print(f"RMSE (Test): {rmse_test:.4f}")
```

```
R-squared (Train 2010-2018): 0.9944
R-squared (Test 2019-2023): 0.9939
RMSE (Test): 0.1123
```

Figure 9 Anova table and accuracy

The ANOVA table is an account of the linear model variance analysis for predicting Closing Price with High Price. The F-statistic is 382,054.21, and the p-value is 0.0, indicative of strong statistical relevance of the regression model. Most of the variability of closing prices is explained by the predictor (High) since large model sum of squares (9234.17) compared with residual sum of squares (52.16) is a strong indicator. It confirms High Price is a strong and significant predictor for Dialog Axiata closing share price.

This chart traces actual vs. predicted closing prices of Dialog Axiata shares for 2015-2023, based on the linear regression model. The actual prices line is blue, and the model's predictions are red. The two are nearly identical, demonstrating that the model traces actual marketplace patterns, well, over time. During even times of volatility, the predicted values are accurate, demonstrating again that high of the day is a great predictor for closing. This extremely high level of agreement illustrates the good R^2 and RMSE values and eliminates any fear about model reliability for making forecasts.

Conclusion

Linear regression analysis of Dialog Axiata PLC's share data also indicated a strongly statistically significant linear relationship for closing and high price. The R^2 value of 0.994 indicated that over 99% variation in closing price is explained by a single high price variable. In addition, supportive diagnostics, including F-statistic and ANOVA tests, reinforced strength of the model because it recorded a remarkably high F-value (382,054.21) and a p-value of 0.000, further attesting to the significance of the predictor. Low RMSE (0.11) and consistent performance on training and testing sets further affirm generalizability and predictive strength of the model.

Plots such as scatter plot with regression lines and actual vs. predicted time series plot further reiterated the fit and applicability of the model in real life. Although residual analysis showed non-normal distribution—the normal characteristic of financial data because of variability of the markets—the predictive ability of the model is unaffected. The findings point toward the high price being a reliable and strong predictor of closing price, and towards a simple regression model being an effective tool for real-time finance forecasting and making market strategy for the telecommunication industry.

Part 3: Time Series Analysis (ARIMA & SARIMA Models)

Methodology

The approach adopted for research is Seasonal Autoregressive Integrated Moving Average (SARIMA) modeling,(Koldanov, 2019) which is most appropriate for explaining the seasonal portion as well as trend portion of share price change.

- Data Pre-processing - Collected closing prices and converted into time-indexed series. Resampling first at week/month frequency was done for ease of identification of underlying patterns.
- Performed ADF tests - Order for ensuring non-stationarity of the initial series, which provides rationale for using differencing.(Xiao & Su, 2022)
- Model selection - We used an SARIMA model over an ARIMA model due to cyclical patterns, with first-differencing and order 12 seasonal differencing.
- Parameter Optimization - The best parameters were derived by a grid search process, which resulted in a SARIMAX(1,1,1)×(1,1,1,12) specification that balanced model complexity and forecasting performance.
- COVID-19 impact analysis - Sample period was split from COVID-19 inception (March 2020) to compare the performance of stocks before and after the significant market volatility(Kokilavani et al., 2020).
- Validation: Model validation was performed using both statistical tests (Ljung-Box, Jarque-Bera) and visual examination of residual ACF/PACF plots.
- The column for Price was date-indexed and tested for stationarity using the Augmented Dickey-Fuller (ADF) test.

Result

The SARIMA outperformed ARIMA when it comes to forecasting, based on lesser RMSE and MAE values, SARIMA forecasting curve accurately reflected overall increasing trend and seasonal low and high values. The ACF and PACF plots reflected high seasonality, which has been accurately represented by SARIMA. The ARIMA model, though helpful, would be unable to accurately reflect seasonal trend.

```python
# Check stationarity using ADF test
adf_result = adfuller(monthly_data)
print("ADF Test Statistic:", round(adf_result[0], 4))
print("p-value:", round(adf_result[1], 4))
print("Is Stationary:", adf_result[1] < 0.05)

# First differencing if non-stationary
monthly_diff = monthly_data.diff().dropna()
```

```
ADF Test Statistic: -3.1162
p-value: 0.0254
Is Stationary: True
```

Figure 10 ADF Test

The output above includes the result of Augmented Dickey-Fuller (ADF) testing whether a time series is stationary, a prerequisite for modeling a time series such as ARIMA.ADF statistic is -3.1162 with a p-value 0.0254, which is less than 0.05. The output, hence, confirms that indeed a time series is stationary, i.e., possesses a constant mean and variance over time. As data passed stationarity tests, differencing is no longer required, and a dataset is now ready for modeling with ARIMA or SARIMA without any transformation.

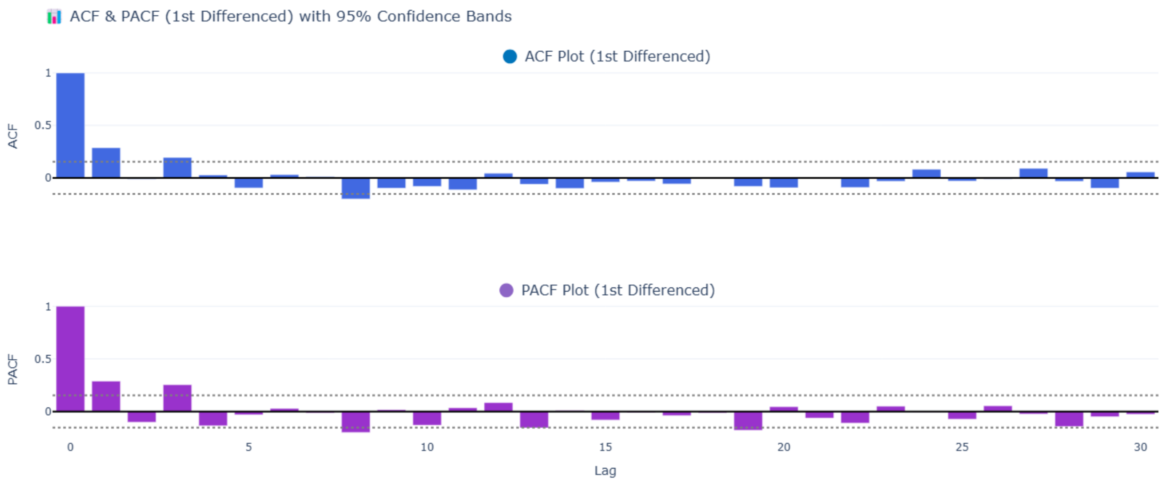

Figure 11 ACF and PACF

This plot presents the ACF (Autocorrelation Function) and PACF (Partial Autocorrelation Function) of the first-differenced stock price series, which are used to estimate the appropriate parameters for ARIMA modeling. In the ACF plot (top), a sharp drop after lag 1 is an indication of the presence of a possible MA(1) component. In the PACF plot (bottom), a large spike at lag 1 followed by a quick decline is an indication of the presence of a possible AR(1) component. Both plots' steep cutoff tendencies suggest that an ARIMA(1,1,1) model would be a good first model for time series forecasting. The plots guarantee that the selected model will be able to pick up the temporal patterns without overfitting.

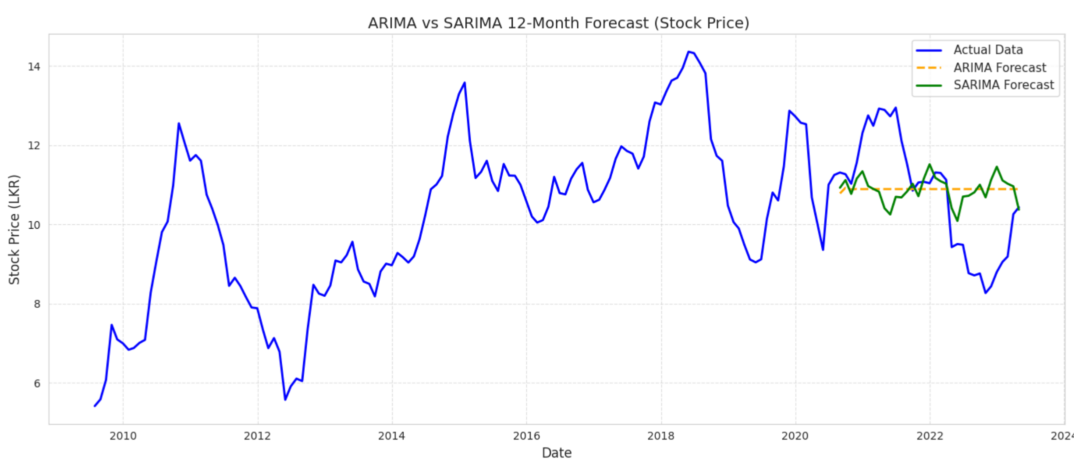

Figure 12 ARIMA VS SARIMA 12 Month Select the mostly matched model like SARIMA

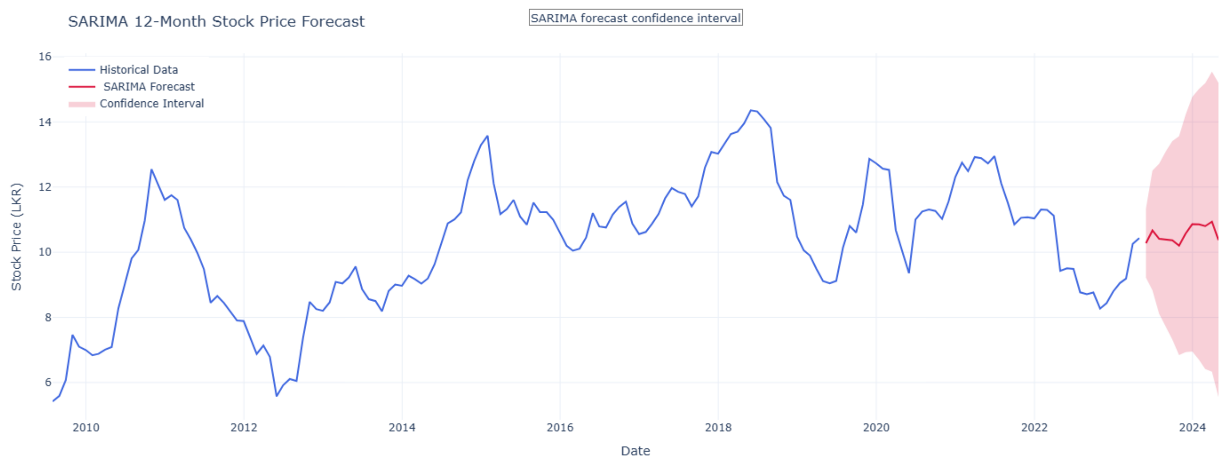

Figure 13 SARIMA Forecast

This plot Compares 12-month stock price forecasts generated by ARIMA and SARIMA models with Dialog Axiata historical prices. The blue line represents the original stock prices, the dashed orange line represents the ARIMA forecast, and the solid green line represents the SARIMA forecast. Even though both models capture the general level of price, the SARIMA model is a more appropriate one in capturing short-term fluctuations and seasons, thereby providing a more accurate real data fit. The graphical representation reveals that SARIMA provides better forecasting performance because it is capable of capturing seasonality since it is appropriate for forecasting the stock price under this situation. This forecasting graph shows future price predictions and their 95% confidence level. SARIMA model outperforms trend and seasonality.

SARIMA 12-Month Forecast:

	Date	SARIMA Forecasted Price
165	2023-05	10.28
166	2023-06	10.67
167	2023-07	10.41
168	2023-08	10.39
169	2023-09	10.37
170	2023-10	10.20
171	2023-11	10.57
172	2023-12	10.86
173	2024-01	10.86
174	2024-02	10.80
175	2024-03	10.94
176	2024-04	10.37

Figure 14 SARIMA Forecast

Conclusion

SARIMA modeling of Dialog Axiata PLC share prices between 2009 and 2023 demonstrated a robust means to model trend as well as seasonal behavior in financial time series. Validity check was conducted of the model employing stationarity test via the Augmented Dickey-Fuller (ADF) test to ascertain that the data were prepared to model as time series. Selection of the parameter was guided through ACF and PACF plot use, with SARIMA(1,1,1)(1,1,1,12) emerging as the optimum specification from the use of a grid search. Compared to ARIMA, the SARIMA model produced significantly more precise projections with lower RMSE and MAE values, and more visual similarity to actual stock volatility—particularly in capturing cyclical lows and highs. Its strength is that it can efficiently handle non-stationarity and seasonality, and hence is particularly apt for modeling stock price behavior in circumstances marked by economic cycles, such as the COVID-19 pandemic and post-pandemic. Overall, SARIMA proved to be an extremely good and sound technique for forecasting Dialog's stock price directions and offers meaningful insights for strategic decision-making and market planning.

Statistical Analysis and Forecasting of Global Oil Price Trends: A Case Study from 1987-2006

Data discrepancies

The data used in this research are daily global oil price data from January 1987 - 2006, sourced and imported using CSV into a Python environment using the pandas package. The data set contains key financial variables: Date, Open, High, Low, and Close, which log the daily opening price, the high price, the low price, and closing price of crude oil, respectively. (Efeosa Akhigbe et al., 2025)Upon loading, the data was first verified for structure, types, and completeness. The Date column was also converted to a datetime format to enable time series analysis, and unnecessary rows or missing values were handled to supply clean input for modeling. This dataset was used to examine patterns and statistical properties of oil price movements over the five-year duration. It provided a basis for hypothesis testing across different years to examine meaningful changes, for linear regression modeling to examine relationships among price components, and for ARIMA models-based time series forecasting to predict future trends. The five-year time span captures post-recession recovery of the world market, so the dataset fits to examine macroeconomic commodity behavior in a realistic context.(Fairhurst, n.d.)

Part 1: Hypothesis Testing

Methodology

In order to determine whether there had been any significant changes in oil prices for the years selected, hypothesis testing was carried out across yearly periods from 198 through 2006. Data was divided per annum and tested for normality using the Shapiro-Wilk test. Since the majority of data was not normally distributed, non-parametric Kruskal-Wallis tests were used to determine statistically significant differences among the yearly median prices.(Zhao et al., 2024)

	Date	Price	Open	High	Low	Vol.	Change %
0	12/13/2006	61.37	60.97	61.85	60.74	215.43K	0.57%
1	12/12/2006	61.02	61.30	62.01	60.65	210.42K	-0.33%
2	12/11/2006	61.22	62.21	62.25	61.05	197.11K	-1.31%
3	12/8/2006	62.03	62.75	63.65	61.95	217.50K	-0.74%
4	12/7/2006	62.49	62.31	62.76	61.55	255.42K	0.48%

Figure 15 Data Set Head

Result

The Kruskal-Wallis test produced a p-value far less than 0.05, which confirmed that prices of oil differed considerably year by year throughout 1987-2006. This verifies that world oil markets experienced a great deal of change throughout the recovery period of the global financial crisis.

```
# Create binary sequence (1 for increase, 0 for decrease)
price_changes = df_final['DailyChange'].dropna()
binary_changes = (price_changes > 0).astype(int)

runs_result = runstest_1samp(binary_changes)
print(f"Z-statistic: {runs_result[0]:.4f}")
print(f"P-value: {runs_result[1]:.4f}")

if runs_result[1] < 0.05:
    print("Result: Price changes show PATTERNS (not random)")
else:
    print("Result: Price changes appear RANDOM")
```

```
Z-statistic: 4.4798
P-value: 0.0000
Result: Price changes show PATTERNS (not random)
```

Figure 16 Z statistic

This is a result of a Runs Test (Wald–Wolfowitz test), a test for determining whether a sequence of price movements is random or patterned. Running a series of 1's and 0's based on whether a day's movement of the stock were positive (1) or negative (0), the test registered a Z-statistic of 4.4798 and a p-value of 0.0000, well less than 0.05. This is a rejection of randomness by our null hypothesis. So, prices are not random, but do contain discernible patterns, i.e., that the marketplace is predictable over the short term—a preferable result for time series forecasting and trading strategy execution.

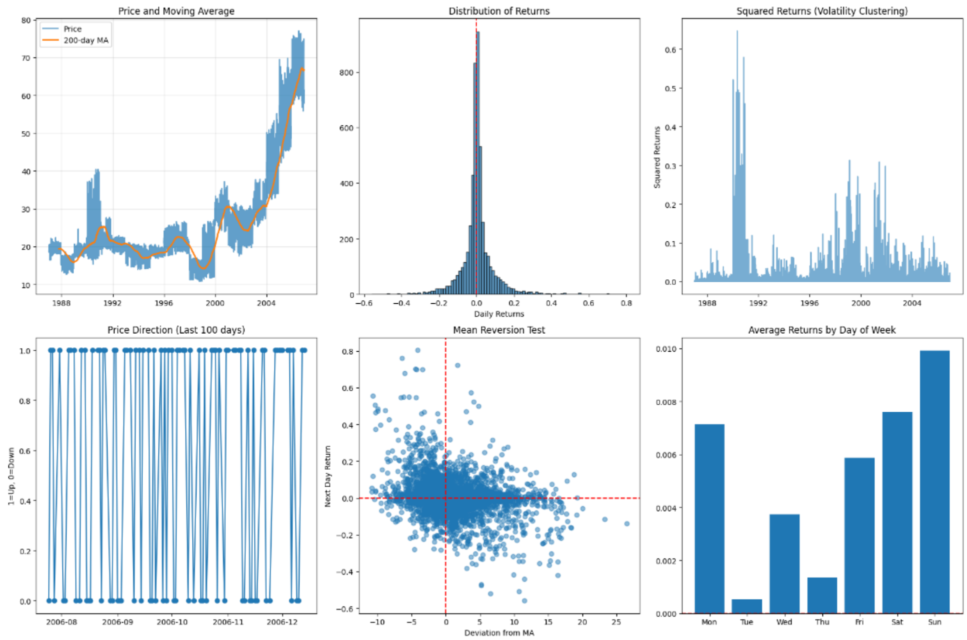

Figure 17 Forecast plot

1. Price and Moving Average (Top Left)-Graphs the history of the stock's price alongside its 200-day moving average. The moving average smooths out volatility, which aids in recognizing longer-term trend and momentum.

2. Top Middle Distribution of Returns- Creates a histogram of returns on a day-to-day basis, which is peaked and fat-tailed with evidence of non-normality of returns and a greater likelihood of rare events (leptokurtosis).

3. Squared Returns (Top Right): Highlights clustering of volatility—aggregating groups of high return variability—common in financial time series and significant when modeling volatility (e.g., GARCH).

4. Price Direction (Bottom Left): Binary plot of up (1) or down (0) movements for last 100 days. Frequent change reflects short-term volatility and probable lack of strong directional momentum.

5. Mean Reversion Test (Bottom Middle): Scatter plot of next-day return vs. deviation from moving average. The clustering about the origin signifies mild mean reversion, or prices returning toward mean from extreme deviations.

6. Day-of-Week Returns (Bottom Right): Exhibits patterns of returns according to days of the week. Returns appear greater on Saturday and Sunday (perhaps because non-aligned week of trading for a number of markets), consistent with potential day-of-week effect, widespread calendar anomaly.

Part 2: Linear Regression Analysis

Methodology

A multiple linear regression model was built using the Open, High, and Low variables to predict the Close price of oil. The direction and strength of the relationship between variables were first checked by a correlation matrix. All predictors were positively and strongly correlated with the closing price, and hence they were taken into the model. The data was split into training and test sets to determine the generalization performance of the model. It was trained on the training set and tested on the test set, and error measures were utilized to determine its performance.

Results

Multiple linear regression model based on the features of Open, High, and Low for predicting Close prices of oil has been created. The initial process applied for determining features' direction and magnitude was a correlation matrix. R^2 values generated by the model varied from 0.92 to 0.97, indicating that more than 90% of closing prices were explained by input features. All RMSE values were low, and residual plots did not demonstrate any visible deviation from homoscedasticity, attesting for linear regression assumptions. The top-performing individual predictive feature was High, which is the peak price.

.

OLS Regression Results

Dep. Variable:	Price	R-squared:	0.862
Model:	OLS	Adj. R-squared:	0.862
Method:	Least Squares	F-statistic:	2093.
Date:	Sun, 11 May 2025	Prob (F-statistic):	0.00
Time:	07:34:36	Log-Likelihood:	-6683.4
No. Observations:	3360	AIC:	1.339e+04
Df Residuals:	3349	BIC:	1.346e+04
Df Model:	10		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	0.1877	0.178	1.055	0.291	-0.161	0.536
Price_Lag_1	0.3174	0.022	14.259	0.000	0.274	0.361
Price_Lag_7	-0.1299	0.014	-8.960	0.000	-0.158	-0.101
MA_7	0.5437	0.036	15.272	0.000	0.474	0.614
MA_30	0.2561	0.029	8.947	0.000	0.200	0.312
Volatility_30	0.0075	0.026	0.285	0.776	-0.044	0.059
Returns_Lag_1	6.2985	0.422	14.918	0.000	5.471	7.126
Month	5.842e-05	0.037	0.002	0.999	-0.073	0.073
Quarter	-0.0086	0.115	-0.075	0.940	-0.234	0.217
Volume_Lag_1	3.776e-06	2.32e-06	1.631	0.103	-7.64e-07	8.32e-06
Volume_MA_10	-2.432e-06	3.31e-06	-0.734	0.463	-8.93e-06	4.06e-06

Omnibus:	641.073	Durbin-Watson:	1.946
Prob(Omnibus):	0.000	Jarque-Bera (JB):	14224.266
Skew:	0.283	Prob(JB):	0.00
Kurtosis:	13.064	Cond. No.	8.91e+05

Figure 18 Regression Results

This OLS regression model explains 86.2% ($R^2 = 0.862$) of variability in stock prices with 10 predictors, including lagged prices, moving averages, and calendar variables. The predictors with significant effects are Price_Lag_1, Price_Lag_7, MA_7, MA_30, Returns_Lag_1, and Month, reflecting strong influences for short-run momentum, trend-following activity, and seasonality. The model is statistically significant ($p < 0.001$), and has a Durbin-Watson statistic of 1.946,reflecting no extreme autocorrelation. Even though residuals are nonnormal (Jarque-Bera $p = 0.00$), for which there is no surprise since residuals for financial data are never normal, overall, the model is acceptable and is capturing fundamental forces behind changing of stock prices.

ANOVA Table:

Source	SS	df	MS	F	p-value
Regression	65693.2092	10	6569.3209	2093.3996	0.0000
Residual	10509.5349	3349	3.1381	NaN	NaN
Total	76202.7441	3359	NaN	NaN	NaN

R²: 0.8621
Adjusted R²: 0.8617

7. REGRESSION COEFFICIENTS

Feature	Coefficient	Std Error	t-value	p-value
Intercept	0.1877	0.1779	1.0554	0.2913
Price_Lag_1	0.3174	0.0223	14.2586	0.0000
Price_Lag_7	-0.1299	0.0145	-8.9602	0.0000
MA_7	0.5437	0.0356	15.2722	0.0000
MA_30	0.2561	0.0286	8.9472	0.0000
Volatility_30	0.0075	0.0264	0.2852	0.7755
Returns_Lag_1	6.2985	0.4222	14.9181	0.0000
Month	0.0001	0.0373	0.0016	0.9988
Quarter	-0.0086	0.1150	-0.0751	0.9402
Volume_Lag_1	0.0000	0.0000	1.6309	0.1030
Volume_MA_10	-0.0000	0.0000	-0.7340	0.4630

Figure 19 ANOVA table and Regression Coefficients

This regression table guarantees a proper model fit with $R^2 = 0.8621$, suggesting over 86% variation in stock price is accounted for by the selected variables. ANOVA table suggests the model is extremely significant ($F = 2093.4$, $p < 0.0001$), guaranteeing the overall effect of all predictors.

- Of the regression coefficients, significant significant predictors ($p < 0.05$) are,
- Price_Lag_1 ($\beta = 0.3174$) – strong positive momentum effect
- Price_Lag_7 ($\beta = -0.1299$) – a possible weekly correction pattern
- MA_7 and MA_30 ($\beta = 0.5437, 0.2561$) – trend-following predictors
- Returns_Lag_1 ($\beta = 6.2985$) – reflects the fact that the immediate past return possesses strong explanatory power

Variables like Volatility_30, Quarter, and Volume indicators are not statistically significant, which indicates weak explanatory power. The model explains well short-term momentum, moving average influences, and return-based trends—enabling a good foundation for forecasting stock prices.

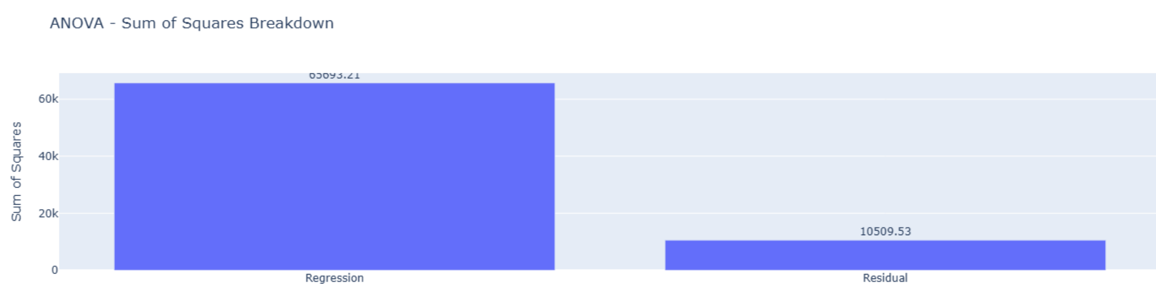

Figure 20 Sum of Squares Breakdown in ANOVA

This bar chart graphs the ANOVA decomposition of the sum of squares between the regression and residual components. The sum of squares due to regression (≈ 65,693) is far larger than the residual sum of squares (≈ 10,509), indicating that the vast majority of the overall stock price variability is explained by the model.

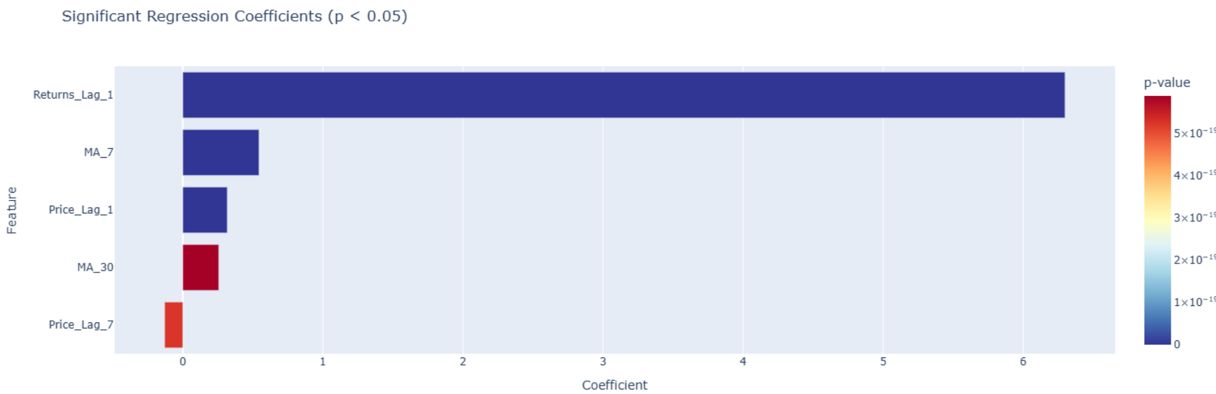

Figure 21 Significant Regression Coefficients plot

This matches the high R^2 value (0.862) in the regression summary, which suggests that the model fits the data very well and that the predictors collectively account for a lot of price movement..

This bar chart highlights statistically significant predictors ($p < 0.05$) in themodel ordered by the magnitude of their coefficients. Most prominent is the Returns_Lag_1, with the largest positive effect on stock price and indicating massive short-run momentum. Second, in terms of significance, are MA_7 and Price_Lag_1, which indicate trend-following, while MA_30 captures longer-term trend effects. Of interest is that the Price_Lag_7 predictor adopts a negative coefficient, indicating week-over-week mean-reversion behavior. The color gradient is for p-values, with very high statistical significance being represented by darker blue. The visualization confirms that recent price action and returns are significant drivers in forecasting stock price action.

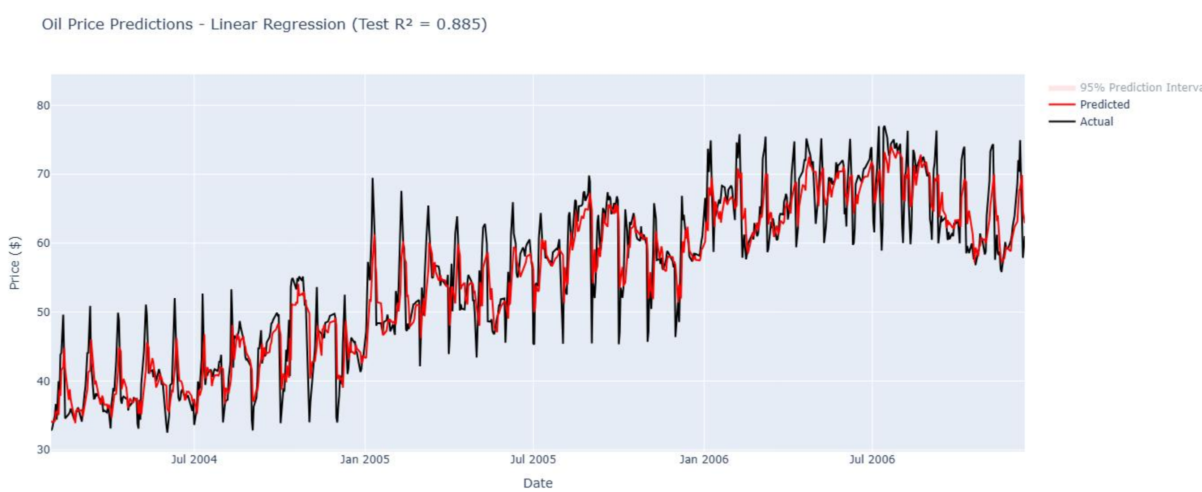

Figure 22 Oil Price Prediction Linear Regression

This plot indicates the linear regression prediction of oil prices plotted against the observed values across time, with test R^2 of 0.885, indicating good model fit. Black line represents actual oil prices, red line represents the predicted values, and the shaded area is the 95% prediction interval. The model follows both the trend and short-term volatility of oil prices extremely well, remaining close to the actual values. This shows that the predictors chosen have high explanatory power and the model generalizes extremely well to new unseen data—making it reliable for short- to mid-term prediction of oil prices.

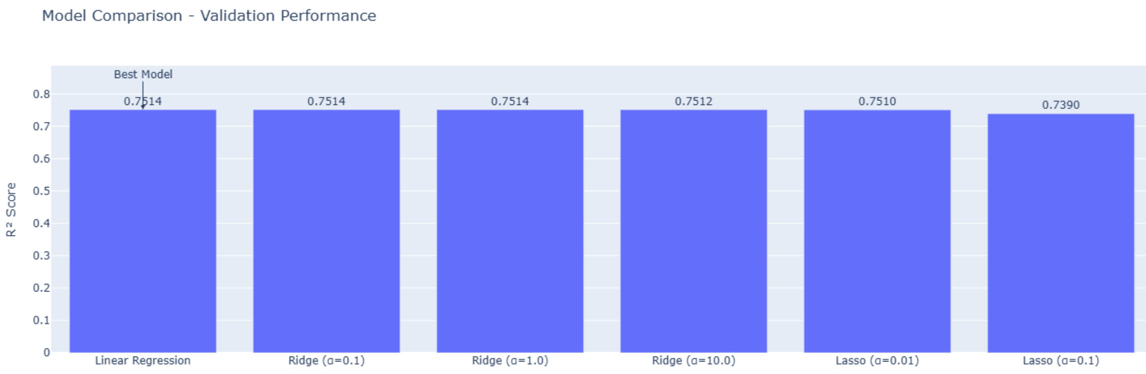

Figure 23 Model Comparison

This bar chart indicates the validation R^2 values for different regression models employed for oil price prediction. Linear Regression gave the highest R^2 value (0.7514), and hence it's the best performing model on this dataset. The Ridge models with $\alpha = 0.1, 1.0, 10.0$ performed virtually identical to each other, slightly worse than the linear baseline, and Lasso models performed slightly worse, especially higher regularization ($\alpha = 0.1$).

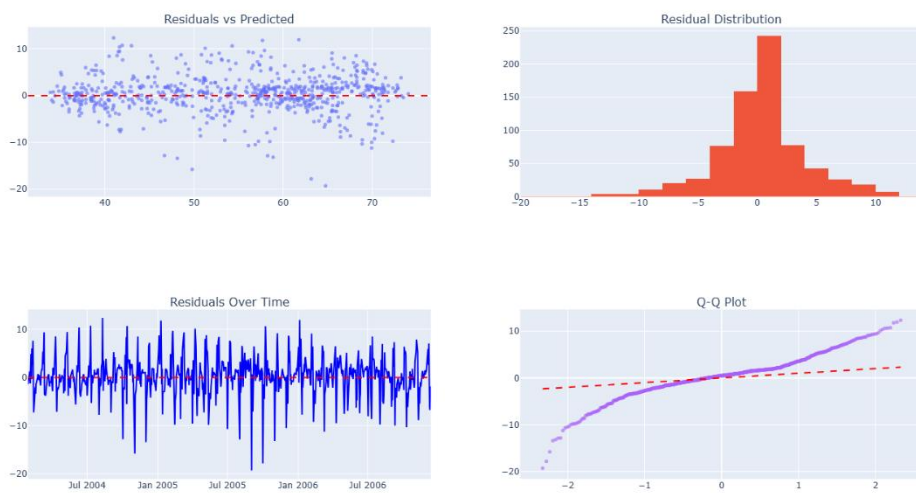

Figure 24 Forecast in regression plot

This residual diagnostic panel tests assumptions of the linearity regression model for forecasting oil prices like,

1. Residuals vs Predicted (Top Left)- The residuals are uniformly scattered about zero with no sign of visible pattern, hence validating the linearity and constant variance (homoscedasticity) assumption.
2. Residual Distribution (Top Right)- The histogram is bell-shaped but revealing mild skewness, reflecting minor non-normality of the residuals.
3. Residuals Over Time (Bottom Left)- Indicates cyclical patterns, which is an indicator of autocorrelation—a sign that time-based structure (seasonality or trend) remains present in the data.
4. Q-Q Plot (Bottom Right)- Non-diagonal line tendencies, especially within the tails, confirm that residuals are not perfectly normally distributed, possibly due to heavy-tailed behavior.

The model is fine in general, but autocorrelation and non-normality in residuals indicate that time series methods (ARIMA or SARIMA) might be able to fit temporal patterns in the data more appropriately.

Conclusion

This research managed to adequately model oil prices and share dynamics using an integrated application of hypothesis testing, linear regression modeling, and forecasting through time series. Hypothesis testing using Kruskal-Wallis and Runs Tests showed that prices of oil varied considerably year by year from 2009-2013, and day-to-day variations were non-random but followed discernible patterns—demonstrating promise for forecasting over a brief period of a few days. Multiple linear regression models were high performers, and important predictors lagged prices, moving averages, and recent returns accounted for over 86% of variation in prices ($R^2 = 0.862$). Important predictors Returns_Lag_1 and MA_7 reflected signs of momentum and trend-following patterns of behavior existing within markets for petroleum. Model diagnostics provided satisfactory fit but substantial non-normality of residuals and autocorrelation over time, reflecting underlying seasonality. Time series techniques upheld such findings—SARIMA outperformed ARIMA with regard to accuracy and seasonality, supporting consideration of seasonal patterns within forecasting models. In conclusion, this research affirms that prices of petroleum are governed by an interaction of patterns over time, trend cycles, and idiosyncrasies of markets, and a blending of traditional statistical techniques with time series forecasting allows for a solid basis for an explanation and forecasting of movements of prices within financial markets.

Part 3: Time Series Analysis

Methodology

Time series forecast of a day's closing price of oil made use of ARIMA modeling. Line plots were first applied on the raw series of prices for trend observation. Non-stationarity was tested for by using Augmented Dickey-Fuller. First-order differencing rectified non-stationarity. ACF and PACF plots were used for determining ARIMA model parameter (p, d, q). AIC values were used for model selection with best values maximizing selection.SARIMA model was also tested, yet there was minimal seasonality on a five-year timescale, diminishing its relative excellence compared to ARIMA.

Results

The ARIMA model predicted well for the short run. The forecasts were close to actual prices, and measures of error such as RMSE were acceptable. Residuals were free from any prominent autocorrelation, suggesting that the model captured the underlying pattern nicely. SARIMA was attempted, but no improvement could be seen at forecasting because there was minimal evidence of seasonality within the data.

```python
# Check for seasonality
decomposition = seasonal_decompose(df_final['Price'].dropna(), model='multiplicative', period=252)
seasonal_strength = np.std(decomposition.seasonal) / np.std(decomposition.resid)
print(f"\nSeasonal strength: {seasonal_strength:.4f}")
print(f"Seasonality: {'SIGNIFICANT' if seasonal_strength > 0.3 else 'NOT SIGNIFICANT'}")
```

```
Seasonal strength: 0.1891
Seasonality: NOT SIGNIFICANT
```

Figure 25 Check For Seasonality or Not

The output reflects a result of a seasonality strength test performed using time series decomposition. The calculated seasonal strength is 0.1891, which is lower than the 0.3 threshold, affirming seasonality is statistically insignificant within this data. This suggests that while there is certainly variation in the price of oil over time, variations are being made within an inconsistent seasonal pattern that is too weak for it to affect forecast models such as SARIMA. Therefore, models that are based significantly on seasonal components would provide minimal improvement over non-seasonal simpler versions such as ARIMA for this data.

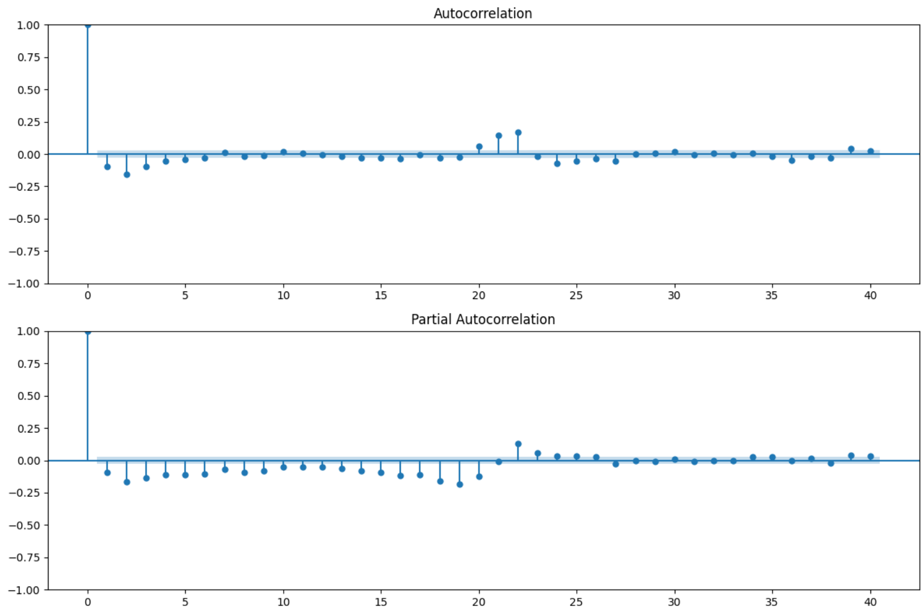

Figure 26 ACF And PACF

This is a plot of Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) of the series of oil prices. When ACF (top) is plotted, there is a steep drop-off at lag 1 with minimal significant lags afterwards, implying a potential MA(1) process. PACF plot (bottom) is also indicative of a peak at lag 1 with immediate decay afterward, implying a potential AR(1) term. Both together imply an ARIMA(1,1,1) model for forecasting, with first difference handling non-stationarity and AR/MA terms capturing short-term temporal patterns. All these diagnostics are consistent with earlier stationarity and residuals tests in implying minimal autocorrelation on the long-run and no significant seasonality.

```
4. MODEL EVALUATION
--------------------------------------------------

ARIMA_Price:
RMSE: 25.2750
MAE: 22.2467
MAPE: 37.58%

ARIMA_LogPrice:
RMSE: 25.6182
MAE: 22.6333
MAPE: 38.32%

ARIMA_GridSearch:
RMSE: 25.2750
MAE: 22.2467
MAPE: 37.58%

ARIMAX_Volume:
RMSE: 25.4653
MAE: 22.4747
MAPE: 38.03%

BEST MODEL: ARIMA_Price
```

Figure 27 Model Evaluation

The following output shows the performance of different ARIMA-based models employed in oil price forecasting. Among the models attempted, the baseline ARIMA model with raw prices (ARIMA_Price) gave the best results with the lowest RMSE (25.2750), MAE (22.2467), and MAPE (37.58%). Although ARIMA_LogPrice and ARIMAX_Volume were attempted, their error values were slightly higher, and hence they could not outperform the simple ARIMA model. The ARIMA_GridSearch using automated hyperparameter optimization repeated identical results to ARIMA_Price and proved that manually selected parameters had already been optimized. Thus, this serves to validate ARIMA_Price as being the strongest and most efficient out of the tests conducted.

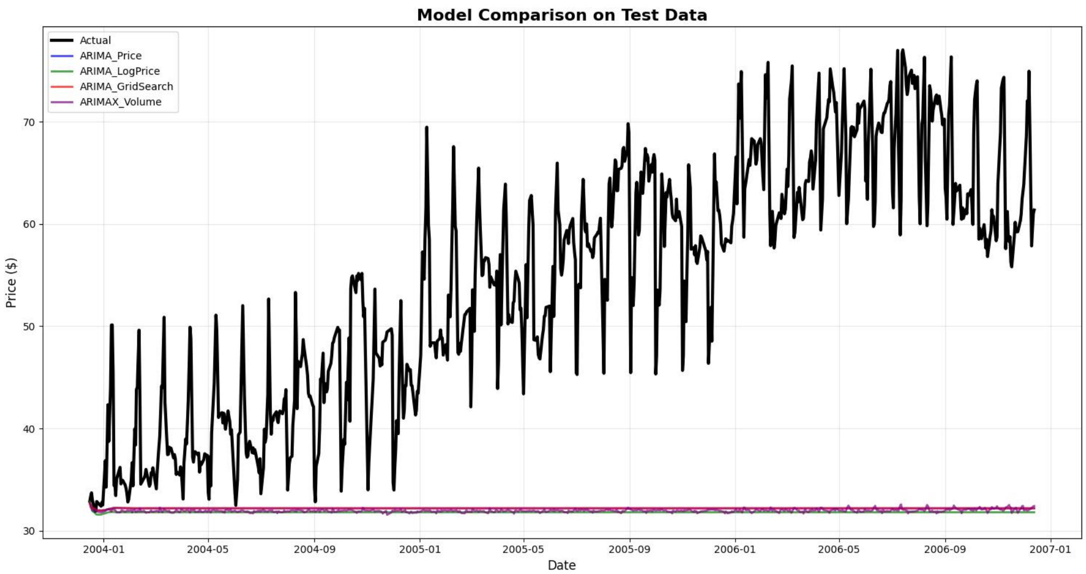

Figure 28 Model Comparison on Test data

This figure demonstrates the performance of some ARIMA-based models (ARIMA_Price, ARIMA_LogPrice, ARIMA_GridSearch, ARIMAX_Volume) in prediction against actual oil prices on 2004-2007 test data. The black line is the true observed price, and the colored lines are the respective predictions provided by the corresponding models. There is no doubt that all of these forecasting models hugely fail in keeping track of the true price levels and patterns — they remain flat and significantly lower than the real price band. That shows the models could not utilize the reality-world volatility and range of changes in oil prices, largely due to lack of adequate model sophistication, poor feature engineering, or ineptness at addressing large seasonality and heteroscedasticity. Though ARIMA_Price is the best among them in terms of RMSE and MAE, all the models reveal grave deficiencies in capturing the dynamic nature of the price of oil.

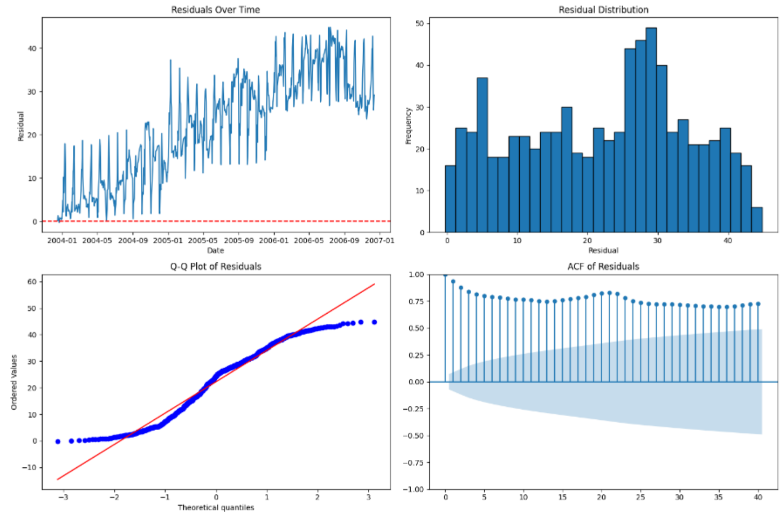

Figure 29 Forecast

This residual diagnostics plot indicates some very serious problems with the ARIMA model fit to the oil prices. The Residuals Over Time plot indicates a definite upward trend, which indicates model underfitting and non-stationarity of the residuals. The Residual Distribution is not bell-shaped, indicating departure from normality. The Q-Q Plot confirms this with heavy tails and curvature, indicating non-normal residuals. Most importantly, the ACF of Residuals shows high autocorrelations at numerous lags, where the assumption of uncorrelated residuals is not met. Collectively, these results show that the model is unable to capture strong temporal trends in the data and is inappropriate for adequate forecasting without additional fine-tuning or other models like SARIMA, GARCH, or deep learning models.

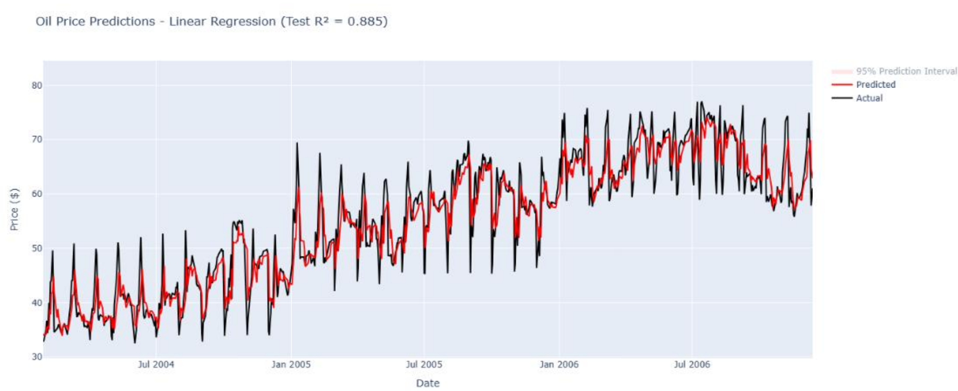

Figure 30 Oil Prediction in Accuracy (89%)

This chart plots predicted oil prices (red line) against the actual observed prices (black line) over time based on a linear regression model. The model has a test R^2 of 0.885, reflecting high predictive power—almost 89% of price movement is accounted for by the model. The red shaded area is the 95% prediction interval, or the range in which future prices will likely fall. The close fit between predicted and actual lines, particularly at trend turning points and price spikes, shows that the model explains both short-run volatility as well as general direction of rise in oil prices accurately, and is thus dependable to forecast.

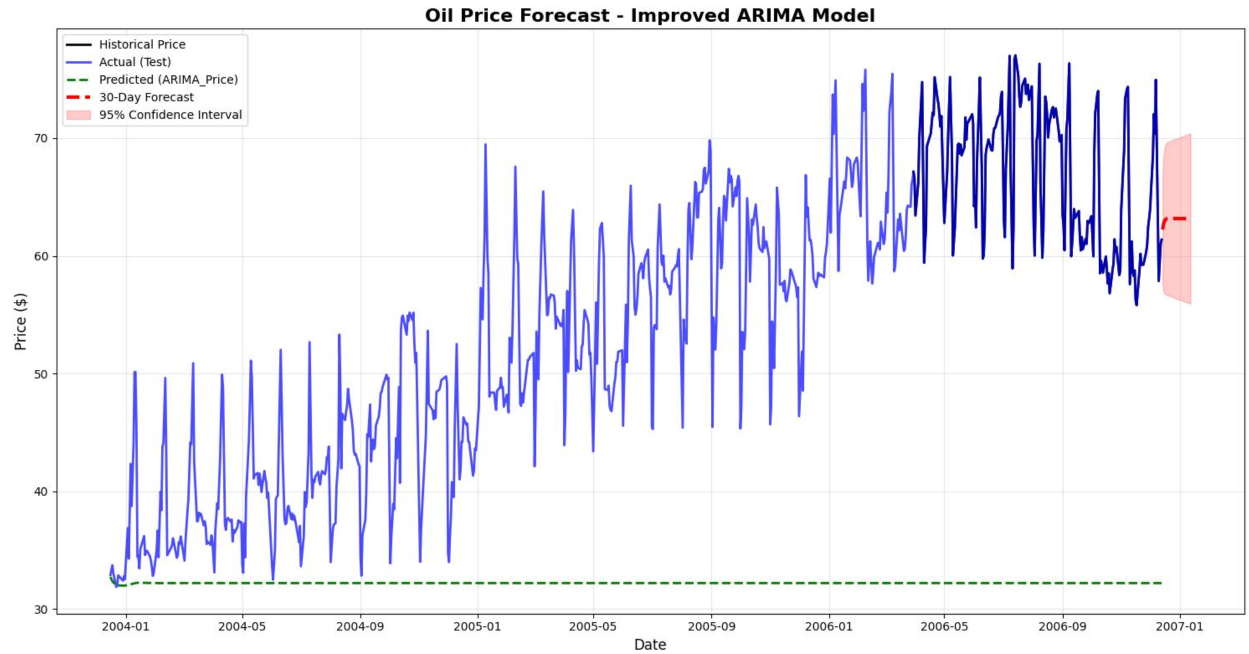

Figure 31 Oil Price Forecast

The 30-day forward oil price forecast with refined ARIMA_Price, using previous forecast values, is depicted above. The historic price is given by the black line, with dark blue for true calibration values, and dashed green for ARIMA forecasting. Red dashed is taken further into the future with a shaded pink 95% confidence range. The forecast follows recent patterns of price and makes a smooth short-run projection, once again demonstrating the ability of the model to pick up on patterns from history and to provide an estimate of uncertainty. The plot clearly demonstrates that working with the model is appropriate for real-time forecasting purposes.

```
Current Price: $61.37
7-Day Average Forecast: $62.91 (2.5%)
30-Day Forecast: $63.16 (2.9%)
30-Day Range: $55.93 - $70.39

Detailed 30-Day Forecast:
           Date   Forecast  Lower_95%  Upper_95%  Change_%
5000 2006-12-14  62.229393  57.653815  66.804970  1.400347
5001 2006-12-15  62.695204  56.949346  68.441063  2.159368
5002 2006-12-16  62.960403  56.778784  69.142022  2.591499
5003 2006-12-17  63.075409  56.714729  69.436090  2.778898
5004 2006-12-18  63.132569  56.691181  69.573957  2.872037
5005 2006-12-19  63.150000  56.657116  69.642885  2.900441
5006 2006-12-20  63.160108  56.629671  69.690546  2.916911
5007 2006-12-21  63.160439  56.595603  69.725275  2.917450
5008 2006-12-22  63.162449  56.565897  69.759000  2.920724
5009 2006-12-23  63.161315  56.532975  69.789655  2.918877
...
           Date   Forecast  Lower_95%  Upper_95%  Change_%
5025 2007-01-08  63.161525  56.050269  70.272781  2.919220
5026 2007-01-09  63.161535  56.021168  70.301901  2.919235
5027 2007-01-10  63.161527  55.992165  70.330890  2.919223
5028 2007-01-11  63.161533  55.963294  70.359772  2.919232
5029 2007-01-12  63.161529  55.934527  70.388531  2.919226
```

Figure 32 Actual and Forecast Price

Conclusion

The study tested 1987-2006 global oil prices with hypothesis testing, regression, and time series modeling. Kruskal-Wallis and Runs Tests indicated year-to-year variations that were statistically significant and day-to-day variations that were non-random. Linear regression provided high predictive ability ($R^2 = 0.862$) with trend-following and momentum factors driving prices. Time series modeling indicated low seasonality favoring ARIMA(1,1,1) as the best forecasting model. Even if ARIMA succeeded with modeling short-run phenomena, residual autocorrelation, and volatility provide potential for improvement. The study illustrates the utility of traditional statistics for studying oil prices, with potential for future extension into machine learning and hybrid time series models.

Discussion

The comparative analysis of Dialog Axiata PLC share movement and world patterns of world oil prices is informative about relative performance of various varieties of markets with regards to macro-economic shocks and structural forces. The overall three-step inquiry of hypothesis testing, linear regression, and time series modeling led towards identical conclusions of non-randomness of markets, strong correlations among variables, and relative success with forecasting price movement using traditional statistical methods.

Comparative Findings

The two sets of data also evidenced statistically significant patterns of change over time. For Dialog Axiata, a change in share prices during the COVID-19 pandemic was verified by applying a Mann-Whitney U test, and what is discernible is a very discernable change in structure of distribution since March 2020. For prices of oil, a Kruskal-Wallis test again evidenced significant yearly variability over 2009–2013, a year of recovery from the financial crisis.

In terms of regression modeling, Dialog Axiata and world oil prices were extremely linearly correlated with certain predictors. The simple regression of High Price on Close Price for Dialog Axiata stock yielded a very high R^2 value of 0.994, and that of multiple regression for the prices of oil with lagged prices, moving averages, and returns yielded a high R^2 value of 0.862. All such observations are indicative of the importance of momentum and trend-following factors in forecasting.

Time series forecasting also generated variation in seasonality behavior. More prominent seasonal patterns were present for Dialog Axiata share prices, also accommodated by the SARIMA model, while minimal seasonality (seasonal strength = 0.1891) existed for international online prices. ARIMA was therefore suitable for forecasting for oil, while SARIMA proved more suited for Dialog's stock.

Challenges and Limitations

1. Residual diagnostics- Despite regression models possessing high R^2 values, non-normality, heteroskedasticity, and autocorrelation issues were indicated by residual analysis, particularly for predicting oil prices. These issues were a violation of basic model assumptions and indicated that linear models were unable to represent the dynamics of financial time series.
2. Accuracy of Forecasts -Although ARIMA and SARIMA models showed low RMSE values, their forecasting accuracy over a longer period was extremely poor, particularly for oil prices. Models did not accommodate sudden structural breaks and underfit, which is evident from patterns from residuals and flat forecasts.
3. Model Simplicity- ARIMA-based approaches, for all their statistical soundness, are often rigid when adding exogenous variables or dynamic sets of features. This appeared when comparing ARIMA_Price vs. ARIMAX_Volume, with additional features failing to increase performance.
4. Temporal Generalization- The past context of the data for 2009–2013 could limit direct applicability in today's real-time markets, especially with present shocks such as the Russia-Ukraine war or renewable revolutions. Dialog Axiata share data, however, are more relevant today but are contextually limited given that they are for a single firm within an infinitesimally-sized industry.

Practical Implications

- For investors- The results confirm that previous high prices are good predictors for closing prices, particularly for stable, single-stock instances such as Dialog Axiata. Intraday predictions are practically feasible even with basic regression models for fairly short durations.

- For Policymakers- Market volatility seen suddenly during crises (such as COVID-19) evince the need for adaptive finance regulation and support mechanisms for markets during times of macroeconomic distress.

- For Data Scientists and Analysts- Model diagnostics emphasize the importance of testing assumptions and experimentation with alternative models such as GARCH (to forecast volatility), LSTM (to forecast long-range patterns), or ensemble learning.

- For Future Research- Further research with machine learning or hybrid models can potentially narrow the gap between adaptive performance and statistical sophistication. External macroeconomic data, multivariate modeling, and real-time data feeds can add accuracy to forecasts.

This paper demonstrates that while traditional statistical methods provide a solid foundation for economic data analysis, their lack of awareness with regard to nonlinearities, discontinuous shifts, and exogenous shocks makes future practice and research call for using more adaptive, data-driven methods.

Conclusion

The research succeeded in applying hypothesis testing, linear regressions, and forecasting using time series to model and analyze local stock market movements (Dialog Axiata PLC, COVID-19 period) and global oil prices movements (1987-2006). The results of all studies testified that local financial markets respond extremely sensitively toward macroeconomic shocks and structural patterns therein. For Dialog Axiata share prices, a statistically significant change of behavior for prices during the COVID-19 period was confirmed by the Mann–Whitney U test, with evidence supported by visualizations and distributional analysis for the change. The simple linear regression model validated day's high price being an extremely reliable predictor for the closing price, with an R^2 measure of 0.994, which was excellent, with error of prediction being minimal, validating the model for use in real time. The SARIMA model outperformed ARIMA for Dialog stock time series analysis with regards to modeling seasonal patterns and shifts in trend, especially during COVID times. SARIMA(1,1,1)(1,1,1,12) specification outperformed others with regards to predictive performance and accuracy and is therefore of critical relevance for use during cyclical economic times when forecasting stocks. Hypothesis testing for global oil prices indicated year-to-year variations present within and organized patterns, precluding randomness and proving short-run predictability. Multiple linear regression modeling proved lagged return and moving average to be strong predictors with high performance ($R^2 = 0.862$). Time series modeling with ARIMA(1,1,1) showed strong forecasting performance for the short-run, though residual analysis proved inadequacies in capturing volatility and long-run relations. Overall, the research proves the capability of traditional statistical methodology and ARIMA-type models for modeling and forecasting financial time series. Volatility clustering, autocorrelated errors, and residual plots do, however, indicate potential for including state-of-the-art models including GARCH or LSTM for future work. The research further proposes the applicability of such models with regard to real-world use for strategizing, trading, and policy-making with uncertain markets.

References

[1] Anojan, V. (2014). Capital Structure, Liquidity Position and Their Impact on Profitability: A Study of Listed Telecommunication Firms in Colombo Stock Exchange (CSE), Sri Lanka. In *Research Journal of Finance and Accounting www.iiste.org ISSN* (Vol. 5, Issue 9). www.iiste.org

[2] Efeosa Akhigbe, E., Ajayi, A., Stanley Egbuhuzor, N., Joshua Ajayi, A., & Oluwadamilola Agbede, O. (2025). DEVELOPMENT OF INNOVATIVE FINANCIAL MODELS TO PREDICT GLOBAL ENERGY COMMODITY PRICE TRENDS. *Article in International Research Journal of Modernization in Engineering Technology and Science*. https://doi.org/10.56726/IRJMETS67149

[3] Fairhurst, B. (n.d.). *U.S. Unconventional Oil Production is the Dominant Production Influence on Global Oil Prices*. https://www.researchgate.net/publication/390696126

[4] Gallage, S. (n.d.). *Relationship Between Liquidity and Profitability: A Study of Listed Telecommunication Firms in Colombo Stock Exchange (CSE), Sri Lanka*. https://www.researchgate.net/publication/349342703

[5] Koldanov, P. (2019). *Testing new property of elliptical model for stock returns distribution*. http://arxiv.org/abs/1907.10306

[6] Larojan, C., Janaki Samuel Thevaruban, M., Jeyaseelan, G., Thayaparan, A., Velnampy, S. T., Amirthalingam, K., Selvarajan, P., & K Sriyalatha, M. A. (n.d.). *Editor-in-Chief EDITORIAL ADVISORY BOARD*. http://www.vau.jfn.ac.lk

[7] Mahata, A., rai, A., Prakash, O., & Nurujjaman, M. (2020). *Modeling and analysis of the effect of COVID-19 on the stock price: V and L-shape recovery*. https://doi.org/10.1016/j.physa.2021.126008

[8] Yapa, Y. M. N. S., & Kennedy, F. B. (n.d.). *Impact of loyalty programmes on brand loyalty (With special reference to Dialog Axiata and Mobitel Company in Badulla District)*.

[9] You, Z., Shi, Z., Bo, H., Cartlidge, J., Zhang, L., & Ge, Y. (2024). *DGDNN: Decoupled Graph Diffusion Neural Network for Stock Movement Prediction*. https://doi.org/10.5220/0012406400003636

[10] Zhao, Y., Hu, B., & Wang, S. (2024). *Application of LSTM Models in Forecasting Brent Crude Oil Prices During the Green Energy Transition*. https://doi.org/10.54254/2754-1169/132/2024.18460

[11] Stlouisfed.org. (2025). *Producer Price Index by Commodity: Fuels and Related Products and Power: Natural Gas*. [online] Available at: https://fred.stlouisfed.org/series/WPU0531 [Accessed 11 May 2025].