# 

# Forecasting Through COVID: Reliance Industries Limited

By:

Bulli Rama Teja Pothula,

Sindhuja Rajasekhar, Benjamin Hood,

Lalitha Valeti, Charles Edwards

Prepared for Professor Sudip Bhattacharjee

University of Connecticut

## Table of Contents

[**Forecasting Through COVID: Reliance Industries Limited**](#_s6tgstrv3ubs) **1**

[Table of Contents](#_tc16osjfshvv) 3

[Executive Summary](#_3odlbf22v5pn) 4

[Introduction](#_jk9w5w2wfk2b) 4

[Problem Description](#_z7wm3pq2jx1a) 4

[Data Description](#_m6gn1i3z943h) 5

[Project Solution](#_voufwxpqsefv) 5

[**Time Series Exploration**](#_tjp9mj579crx) **6**

[Stationarity: Dickey-Fuller Test](#_h78e0askkoqd) 9

[Pre-Whitening & Cross-correlation](#_rupcn85on57) 10

[Variable Selection](#_7dereis7ahrc) 12

[**Forecasting**](#_6s18326hlncp) **13**

[Exponential Smoothing](#_9tam1bfulxq5) 13

[Cross correlation analysis for choosing the model:](#_7f9dkcaucngw) 15

[ARIMA:](#_ydz2lfeo7kao) 16

[ARIMAX](#_iipy133m1ik9) 20

[Model Comparison](#_afzkj2jn8wu2) 22

[**Conclusions**](#_cbz7secpdrvn) **24**

[Recommendations](#_cwrr2ceif6cu) 24

[**Citations**](#_dmjy5c5cdg97) **26**

### Executive Summary

The COVID-19 pandemic has been an international event which has disrupted daily life in many ways. Risk and uncertainty in the financial sector are always issues for investors, but how much does an event like the pandemic affect our ability to model the financial market? Are there safe stocks which can be modeled and whose models continue to work in a pandemic?

Our hypothesis is that a large commodity stock’s forecasting model would hold through the pandemic and would have the most resilience. We chose Reliance Industries because it is a petrochemical conglomerate and the largest private sector corporation in India.

This white paper explores the stock price of Reliance Industries as time-series data, develops several forecasting models, and evaluates those models using stock values during the COVID-19 pandemic. The most effective model will be used to determine if COVID-19 was indeed an ‘event’ that would disrupt our ability to forecast stock price. We conclude with a series of recommendations based on the outcome and our observations during the project.

### Introduction

#### Problem Description

Economic markets around the world struggled in the wake of the COVID-19 pandemic, which firmly took hold in March of 2020. Our hypothesis was that commodities would be the least vulnerable to fluctuations in price and, therefore, a forecast model of the data prior to the pandemic would remain relevant during the pandemic.

We looked at India’s NIFTY 50 index for commodities to study. We attempted to combine a number of stocks in a similar commodity for comparison, but this proved non-trivial within SAS Studio. We then settled on Reliance Industries Limited (RIL), a conglomerate of energy, materials, textiles and retail brands, for our study.

#### Data Description

We took our data from a Kaggle dataset of India’s NIFTY 50. India's National Stock Exchange (NSE) provided the source data which spans from January 1, 2000, through April 30, 2021. From Kaggle alone this data set has been downloaded 14 thousand times and Kaggle users have used it to develop trading algorithms. (Vopani)

NIFTY 50 is a stock index of the 50 largest companies on the NSE. Industries included in the index are banking and financial services, pharmaceuticals, energy, automotive, information technology, consumer goods, and commodities. The dataset classifies the companies in 13 industry sectors.

RIL is the largest private sector corporation in India. They began as a textile and polyester company and is now a large petrochemical conglomerate. (*Fortune 500 India*) RIL has responded to the COVID-19 pandemic by ramping up the production of medical grade oxygen to over 1000 metric tons daily, and has produced over 55,000 metric tons since March of 2020. RIL now supplies 11% of India’s medical grade liquid oxygen. (Reliance)

#### Project Solution

In order to prove our hypothesis, the RIL dataset was broken into a pre-January 2020 dataset for training and validation and a post-January 2020 dataset for forecasting. The post-January 2020 dataset represents the COVID-19 pandemic data. Several forecasting models were built using the training and validation datasets and the forecasts evaluated on the COVID-19 data. If the forecasting model accuracy is unchanged from the validation set (2019) to the COVID-19 data, we can conclude that the pre-COVID forecasting model still works and that RIL stock prices were reasonably predictable throughout the pandemic.

### Time Series Exploration

In this step, we analyzed the data by looking at trend, seasonality, stationarity and correlations between variables, and understanding its effect on the forecasting. For this, we first considered Close as a dependent variable and rest all other variables as independent variables to understand it’s auto-correlation and cross correlation.

The 2000-2020 RIL close data in raw form is adjusted to weekly time series intervals, this was adjusted because when considering the data day wise, due to high volume of data points the 20 years of data was not displayed accurately in the graph. Here we considered weekly average data. Between 2000 and 2008, stock prices grew by an order of magnitude.



Trend

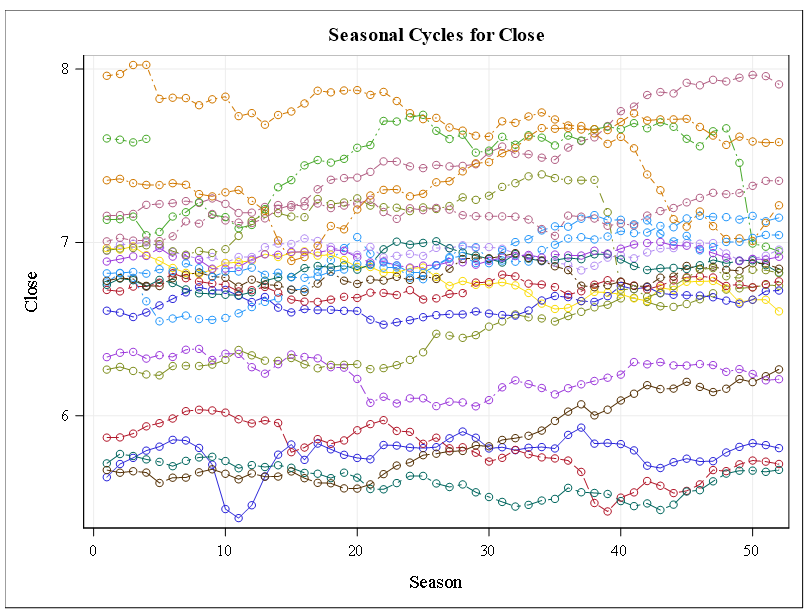


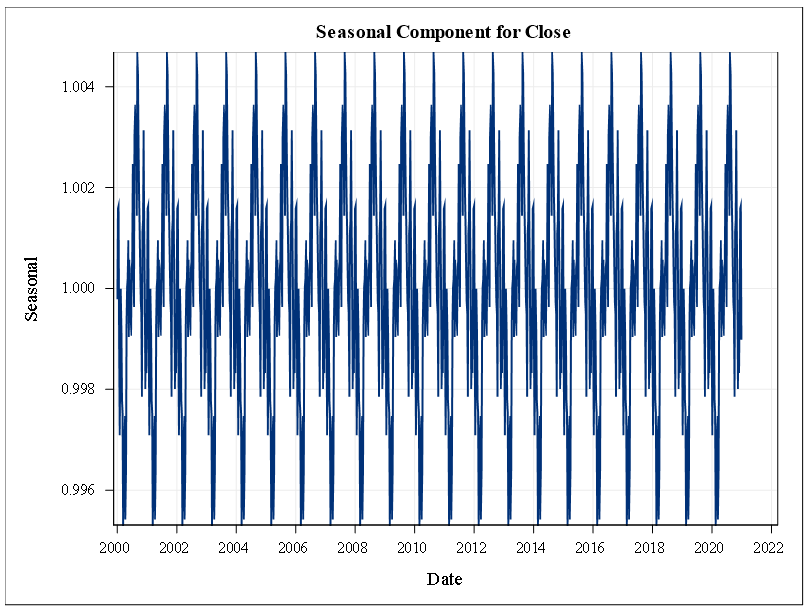
Looking at the RIL stock closing price through 2020, there are several nearly linear trends. From 2003 to 2008, RIL stock prices grew by nearly 50%. This trend was ended by the global recession in 2008. RIL closing prices took on a downward trend from 2008 to 2012. From 2012 to 2020, RIL began another steady upward trend, but did not match the peak of late 2007. There is a trend in this data, even though it has visible spikes at certain points.

Certain events that have influenced the spikes are as follows:

* 2008 - Global recession
* 2010 - Acquisition of broadband services- Infotel, successful bidder for pan india 4G spectrum auction.
* 2017 - Jio venture
* 2018 - Comments of the Finance minister's proposal in the budget speech

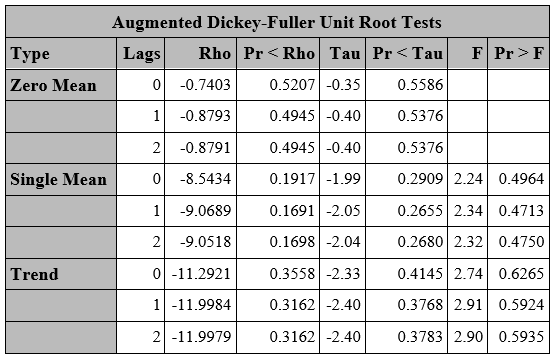
**Seasonality**



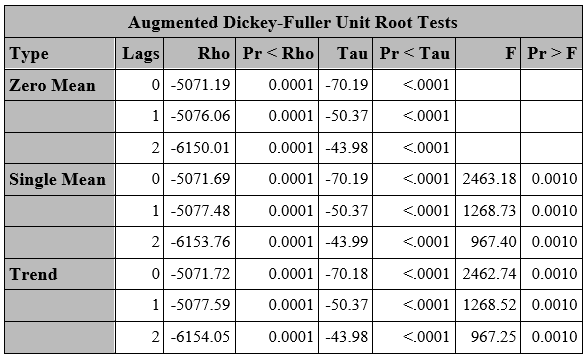


There is no apparent seasonality in either the original or the differenced seasonal cycles plot. Being a large conglomerate, RIL is likely shielded from many seasonal fluctuations that may affect corporations in a single industry.

#### Stationarity: Dickey-Fuller Test



The Dickey-Fuller Unit Root Tests of the original data have p-values much greater than 0.05. This indicates that the time series data is not stationary and therefore cannot be modeled with an ARMA model.

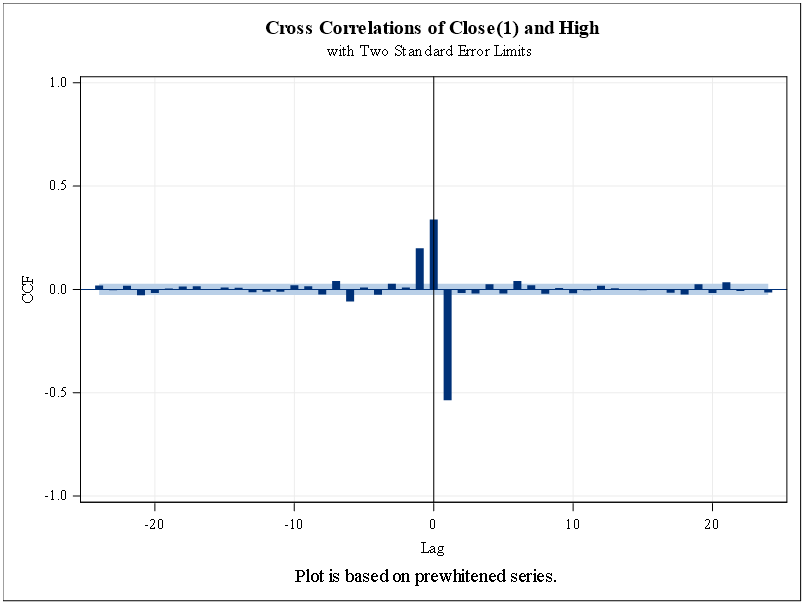
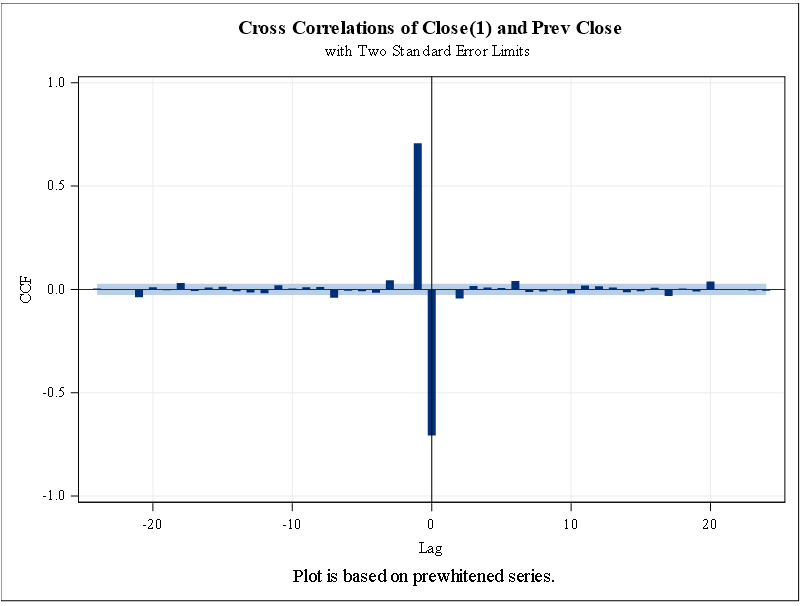


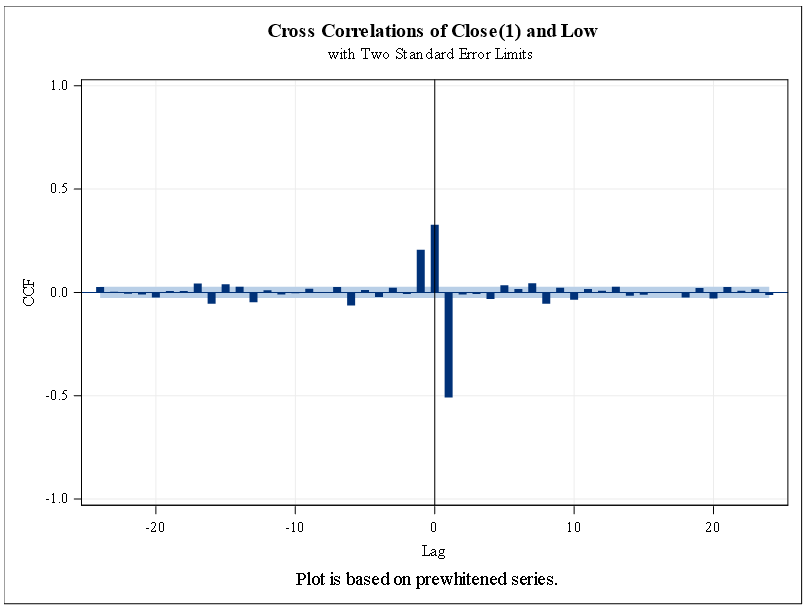
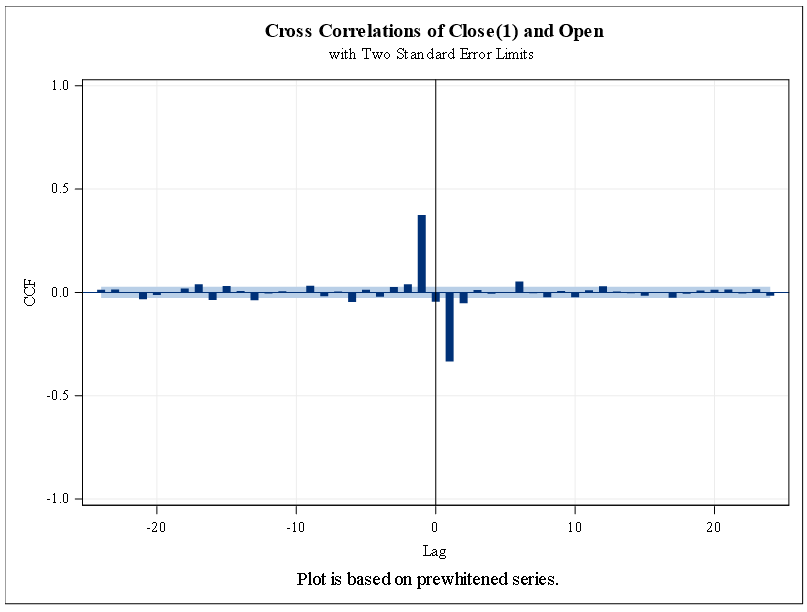
After differencing an order of 1, the Dickey-Fuller Unit Root Tests confirm that the time-series is stationary. Hence we can use this time series for further modeling, ARIMA model can be used with a differencing order of 1.

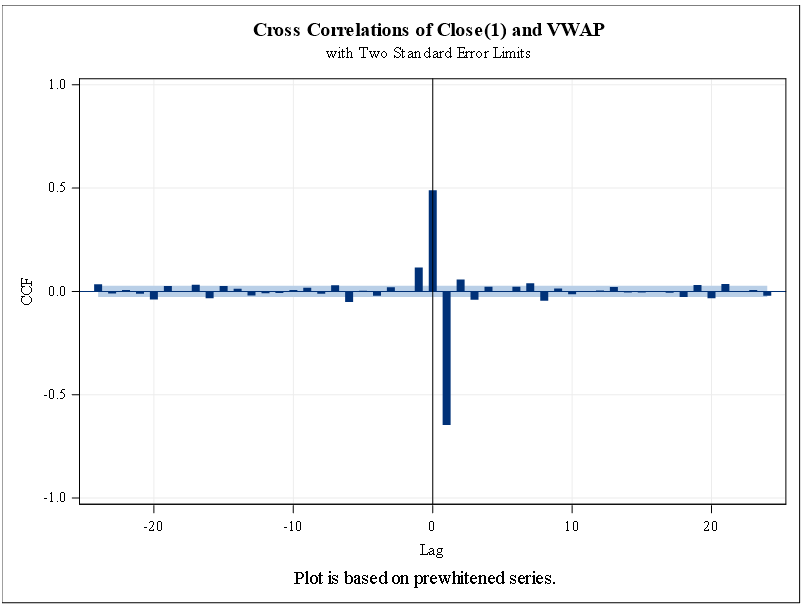
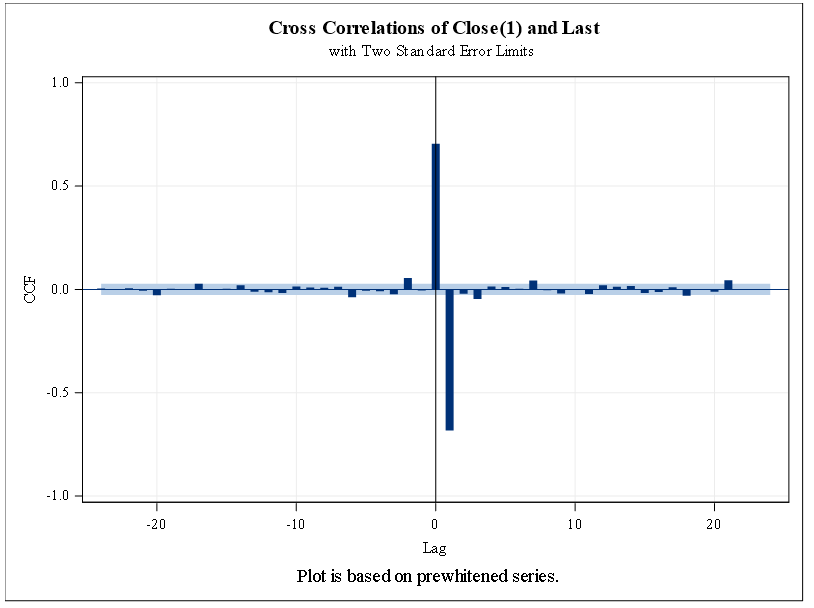
#### Pre-Whitening & Cross-correlation

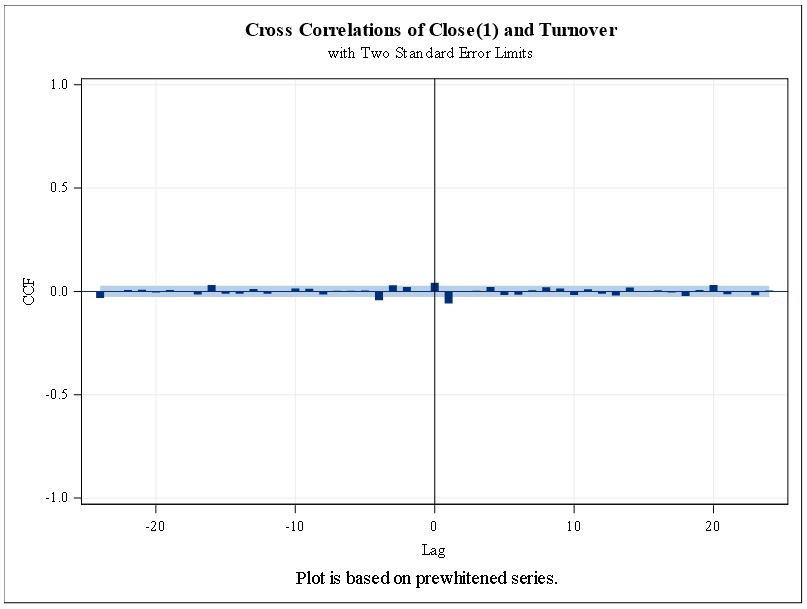
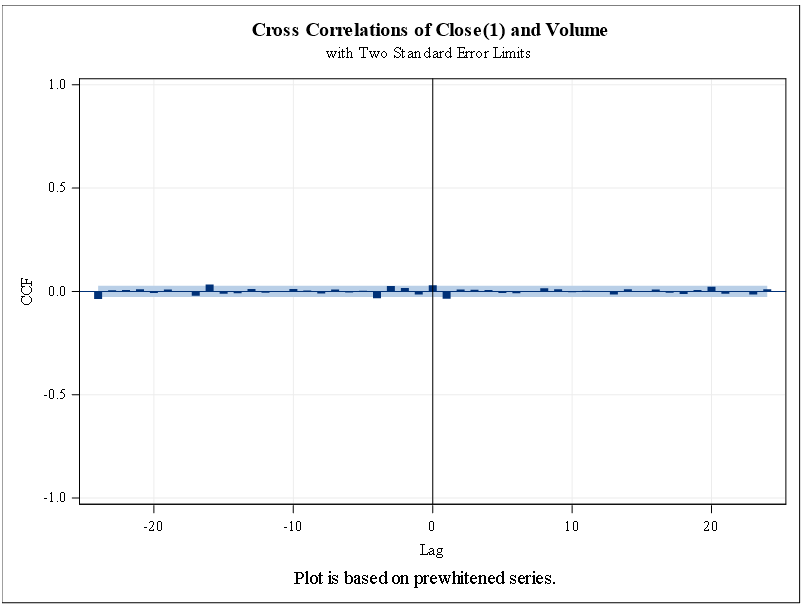
The cross-correlation function (CCF) assists us figure out which time series X lags predict the value of time series Y. However, identifying meaningful links between the two time series is difficult if either series has autocorrelation or the two series have common patterns. Pre-whitening eliminates autocorrelation and trends, which solves the problem. Hence, we performed pre-whitening to confirm if the variables have cross-correlation.

After pre-whitening we observed that the variables Previous close, High, Open, Low, Last, VWAP, Volume and Turnover have correlation with the dependent variable. Among these variables we can see that Previous close, High, Open, Low, Last, VWAP has high correlation; this can be confirmed by looking at the CCF values at different lags.









#### Variable Selection

Initially, the team was investigating VWAP as the dependent variable. VWAP, or volume weighted average price, unfortunately is already a weighted average value and is a modeled parameter. Domain knowledge would have avoided some wasted effort on our part.

The daily closing price is the final stock price for the day and therefore is subject to all of the possible variations both internal and external to RIL, the stock market, and the market in general. We chose the closing variable for our dependent variable.

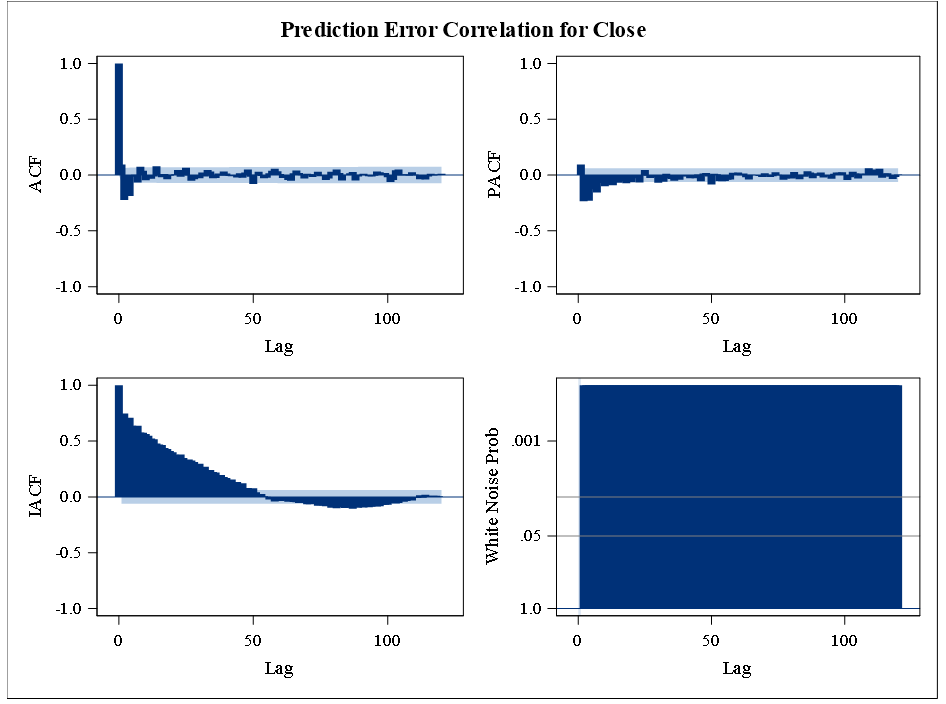
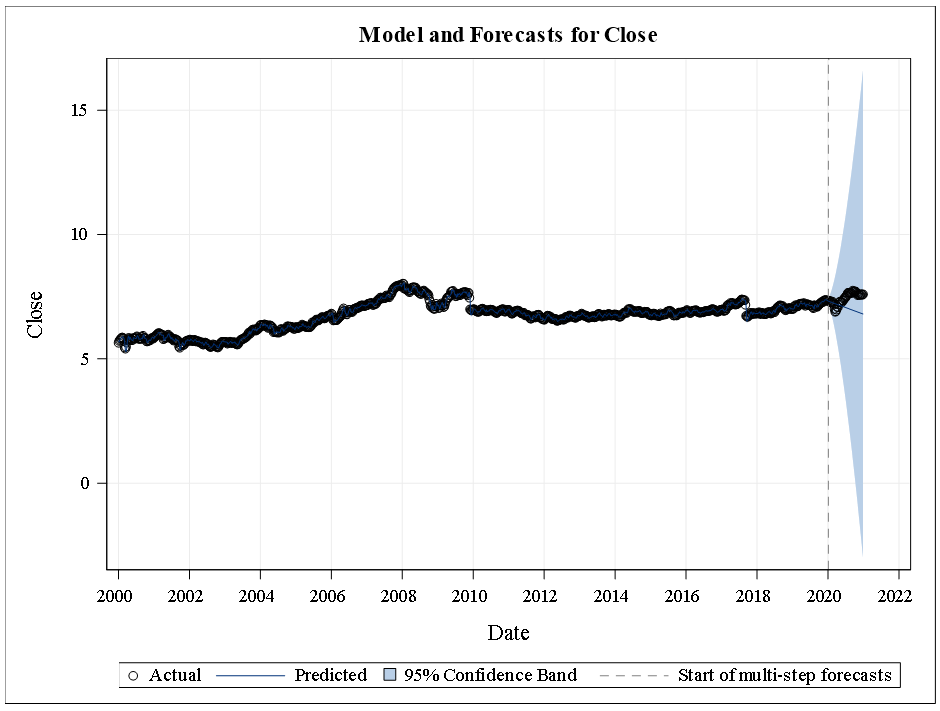
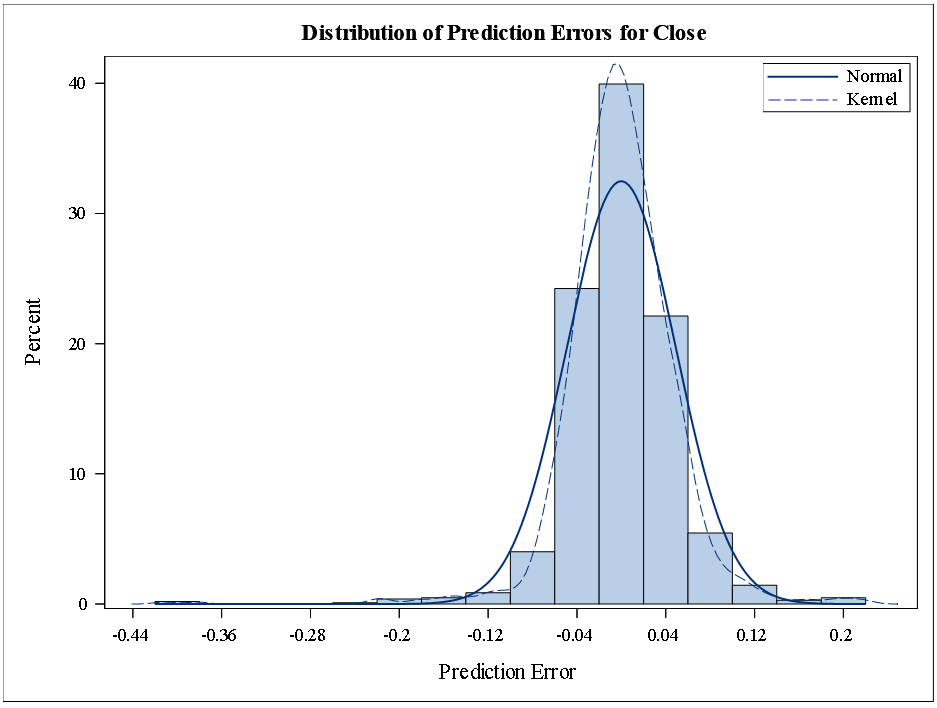
The variables available in the data set include VWAP, close, open, high, low, previous close, volume, turnover, and date. However, we picked Previous close as the independent variable, as the rest of the variables are from the same day as the dependent variable, hence it does not make it a practical selection. We used this independent variable in the model building of the ARIMAX model.

### Forecasting

#### Exponential Smoothing

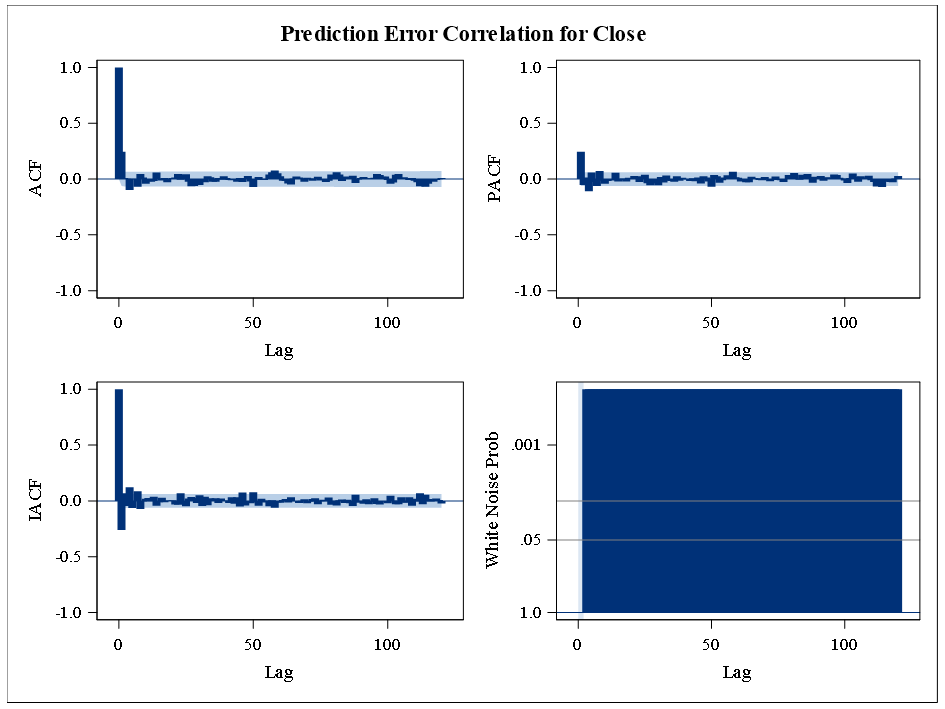
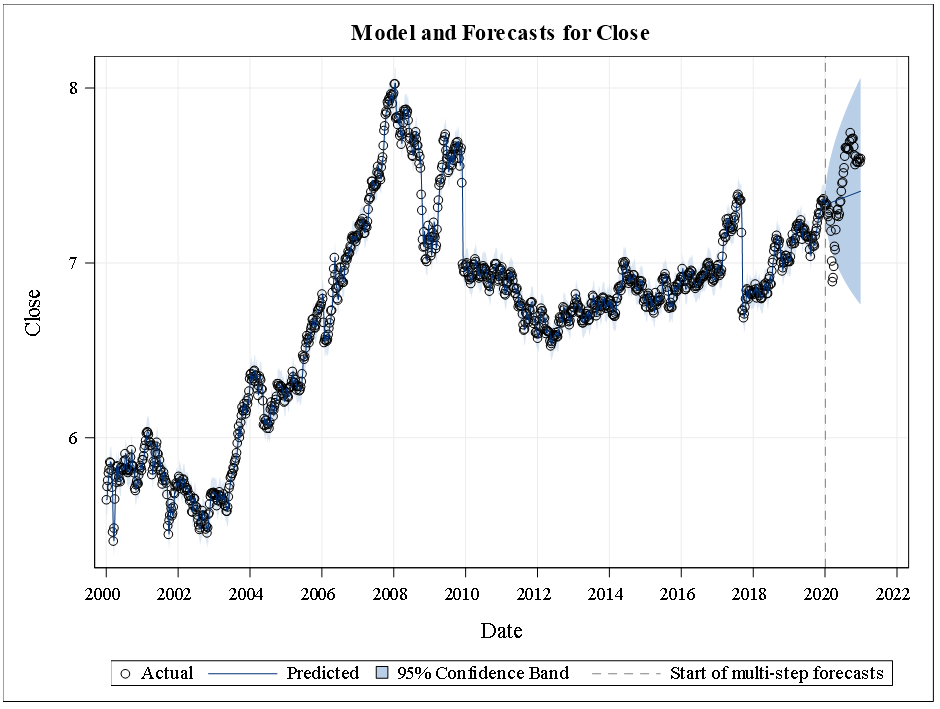
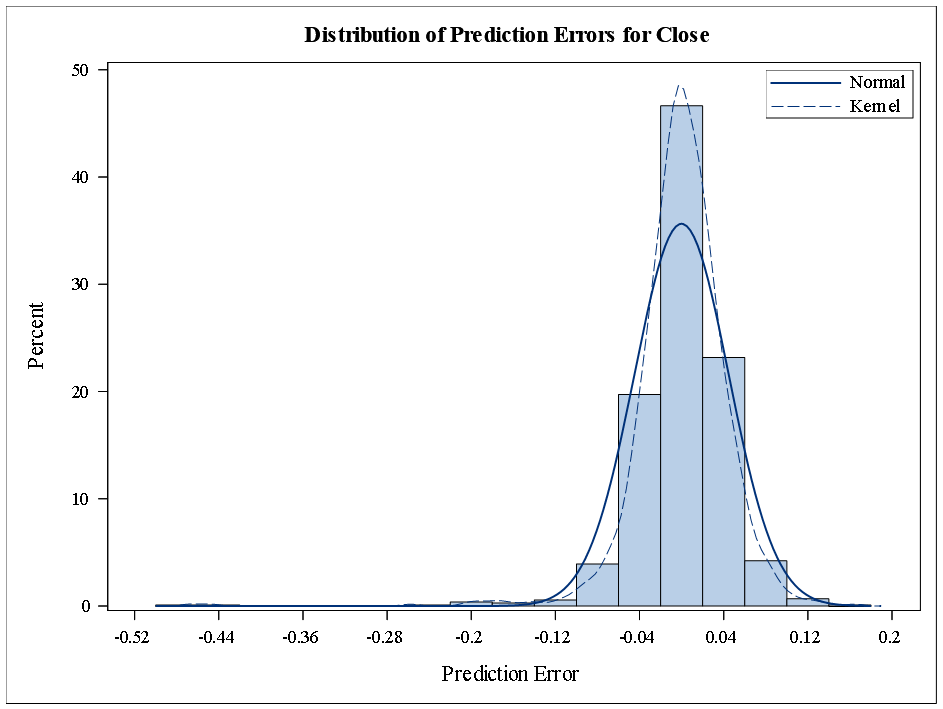
The RIL data shows no seasonality but does have some trend, so we will have Double Brown Exponential Smoothing models, Holt models, and a DAMP model.

Brown Model:



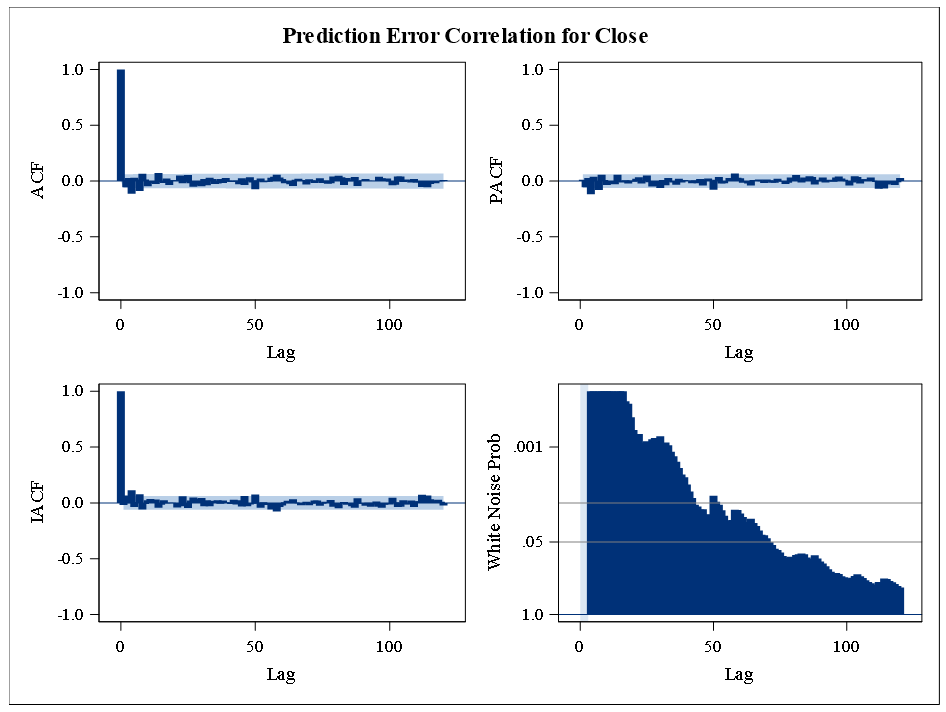
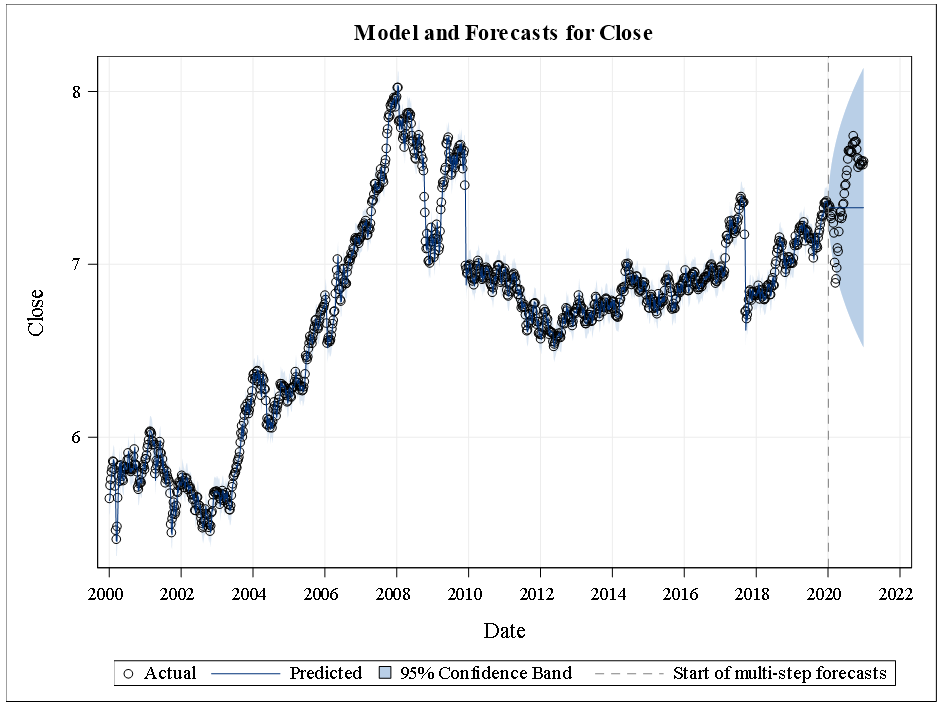
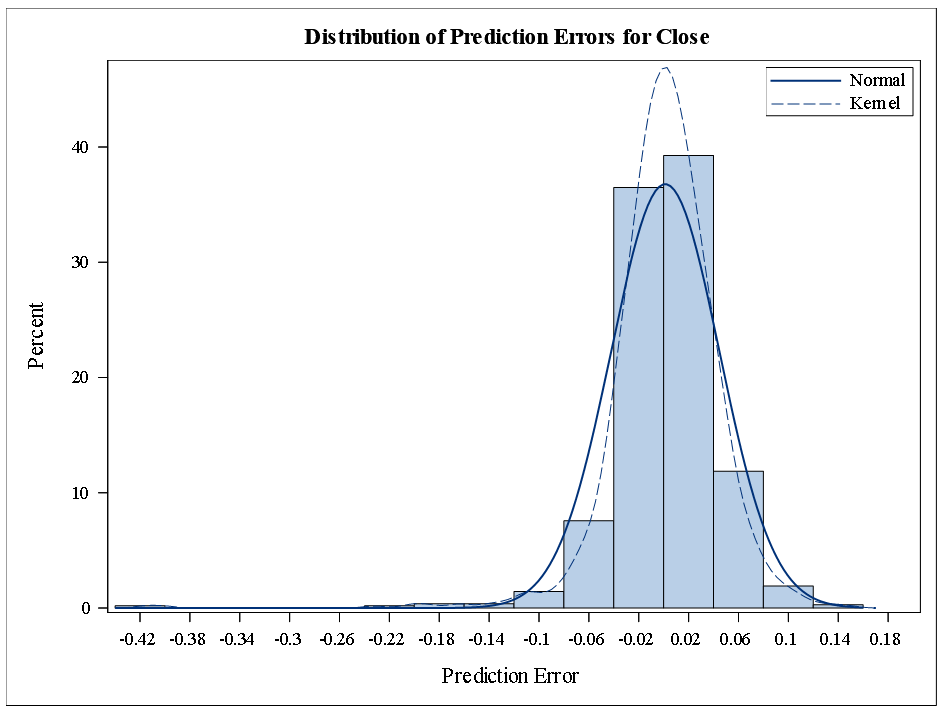
The errors seem normally distributed shown in the histogram with the superimposed normal and kernel curve. According to the white noise probabilities there do seem to be attributes that can be modeled within this time series. However looking at the ACF there is no autocorrelation. Also the forecast does a poor job of modeling actual close prices.

HOLT Model:



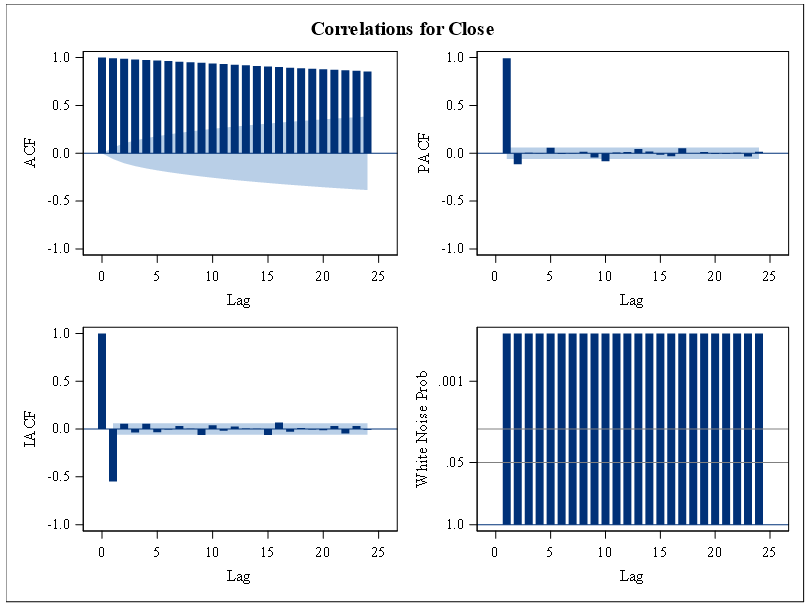
The errors seem normally distributed shown in the histogram with the superimposed normal and kernel curve. According to the white noise probabilities there do seem to be attributes that can be modeled within this time series. However looking at the ACF there is no autocorrelation. Also the forecast does a poor job of modeling actual close prices, but it does model the positive trend.

DAMP Model:



The errors seem normally distributed shown in the histogram with the superimposed normal and kernel curve. According to the white noise probabilities there do not seem to be attributes that can be modeled within this time series model. However looking at the ACF there is no autocorrelation. Also the forecast does a poor job of modeling actual close prices.

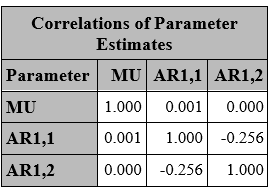
#### Cross correlation analysis for choosing the model:

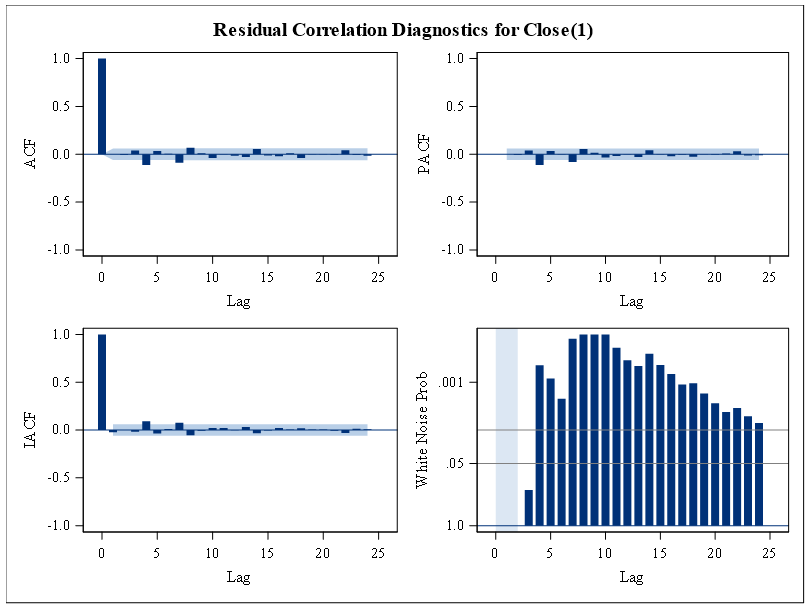


The raw data is non-stationary based on the Dickey-Fuller results (above), so a pure ARMA model is not appropriate. With a differencing order of 1, the Dickey-Fuller results indicate that the transformed data is stationary. Also, from the ACF, PACF, IACF plots we can choose the AR model to build on our data, as the ACF plot shows a decrease trend in the lag values and the lags in the PACF, IACF are abrupt at 1,2 lags. Therefore, an ARIMA or ARIMAX model would be appropriate.

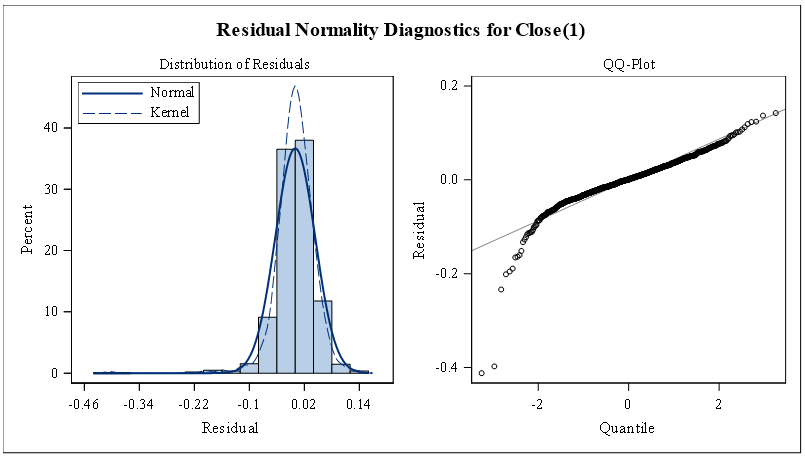
#### ARIMA

**Differencing = 1, Number of lags = AR = p= 2**

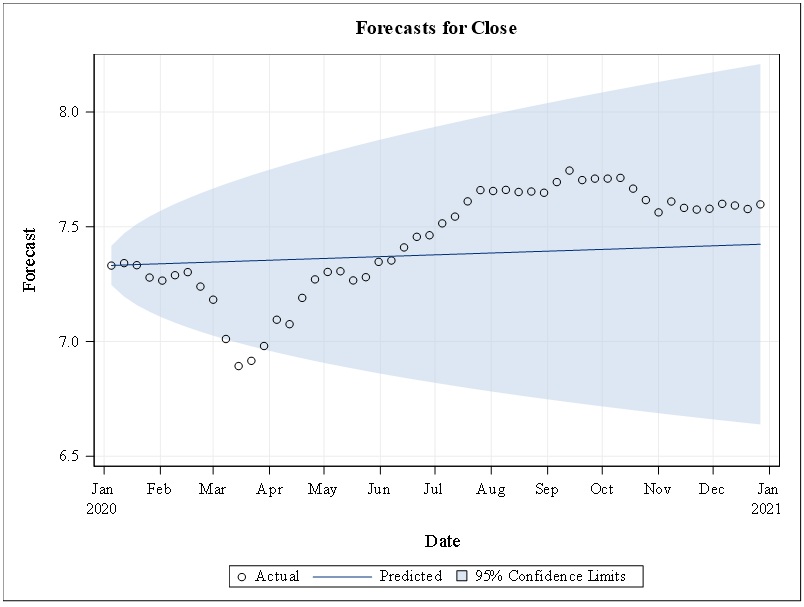




From the above residual plots we can see that there is no significant correlation between the residuals across the first 24 lags. There is very little white noise.



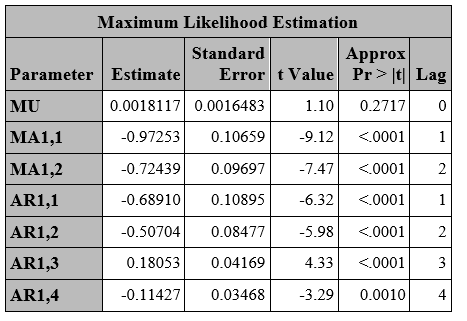
As the residuals are normally distributed we can say that our model fits well.

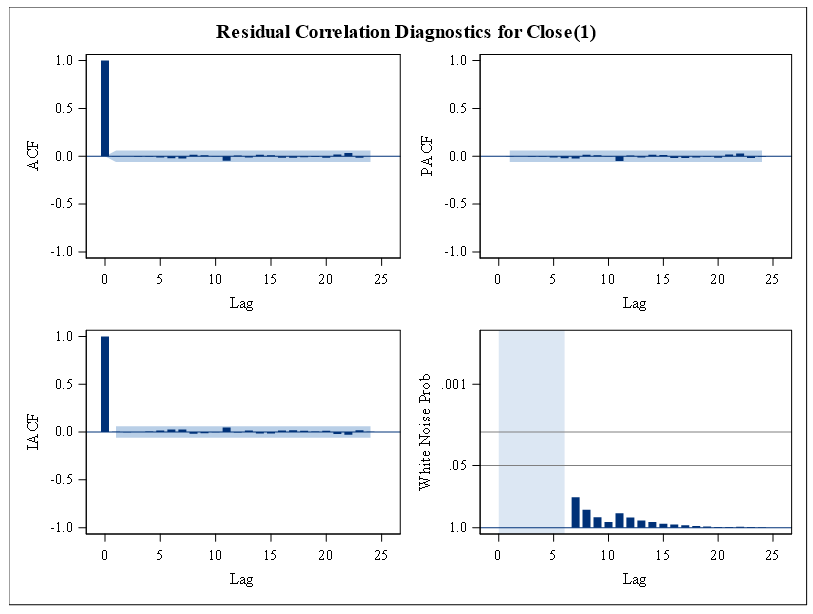


From the above forecast plot we can say that all our predicted values are present within the 95% confidence interval.

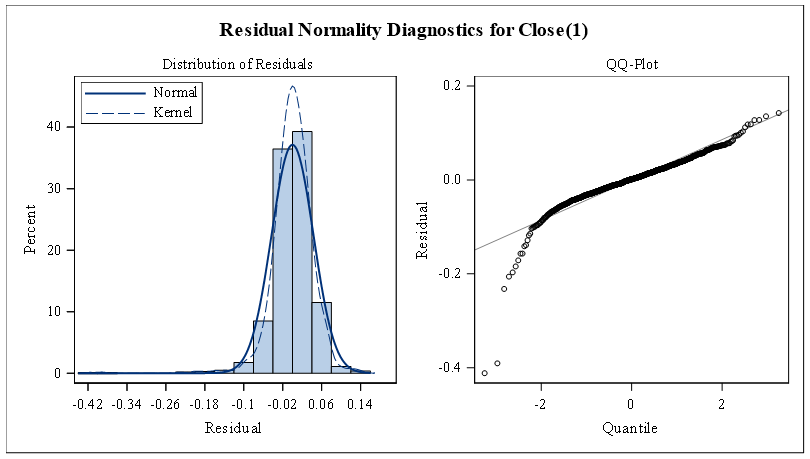
As there is less whitenoise, we tried to play with the p and q values in order to reduce the residual error and increase the white noise. By doing so we found that the below model did well.

**Differencing = 1, Number of lags = AR (p)=4, MA (q)=2**

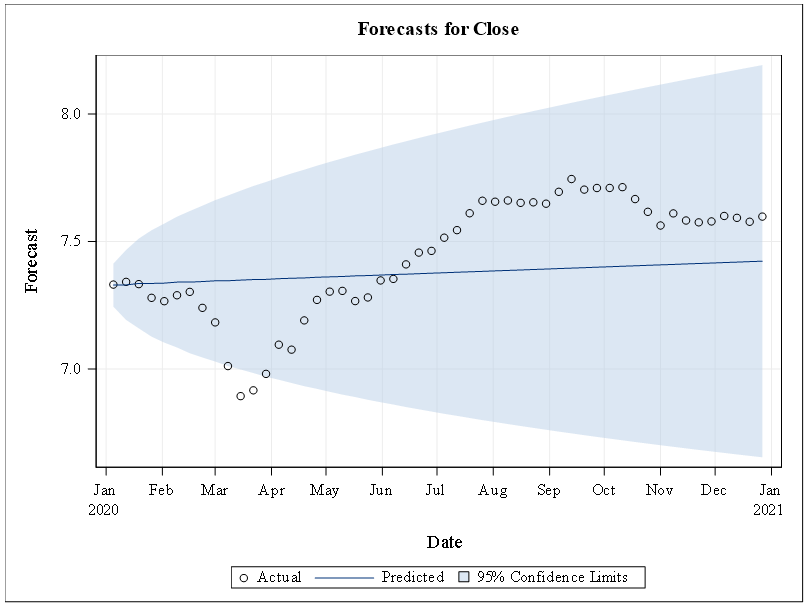




From the above residual plots we can see that there is no significant correlation between the residuals across the first 24 lags. There is a lot of white noise.



As the residuals are normally distributed we can say that our model fits well.



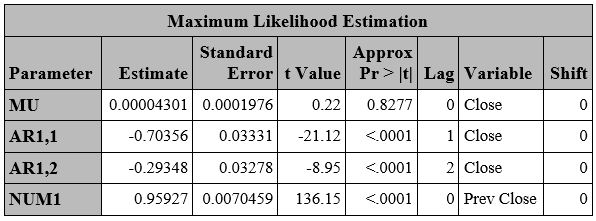
From the above forecast plot we can say that all our predicted values are present within the 95% confidence interval.

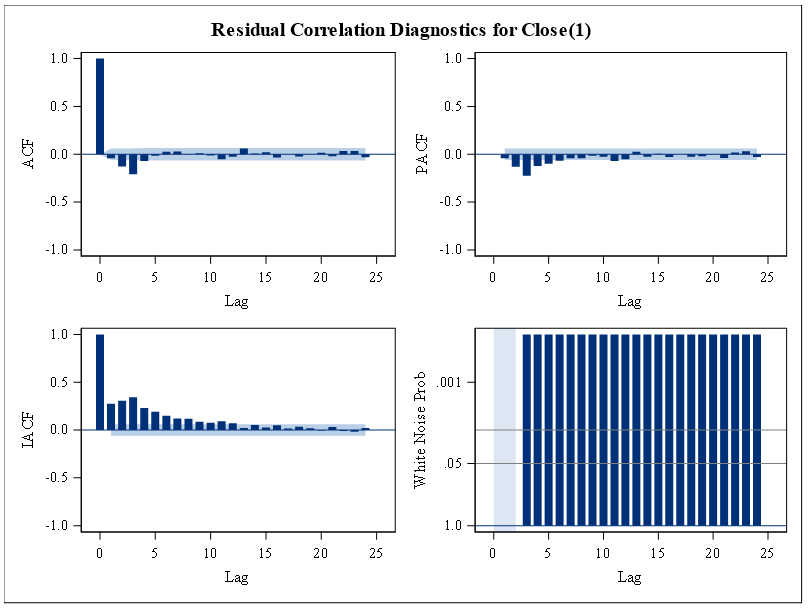
#### ARIMAX

**Differencing = 1, Number of lags = AR (p)=2**

**Additional variables = Previous Close.**

As discussed in variable selection, here we built a model with only one independent variable i.e., previous close in order to forecast the Close value.



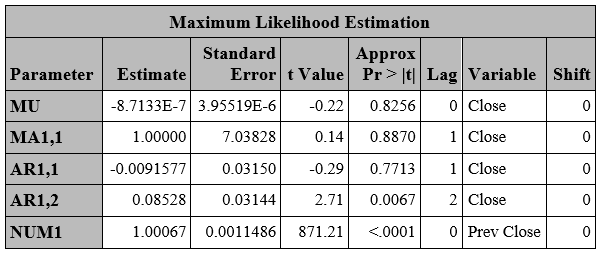


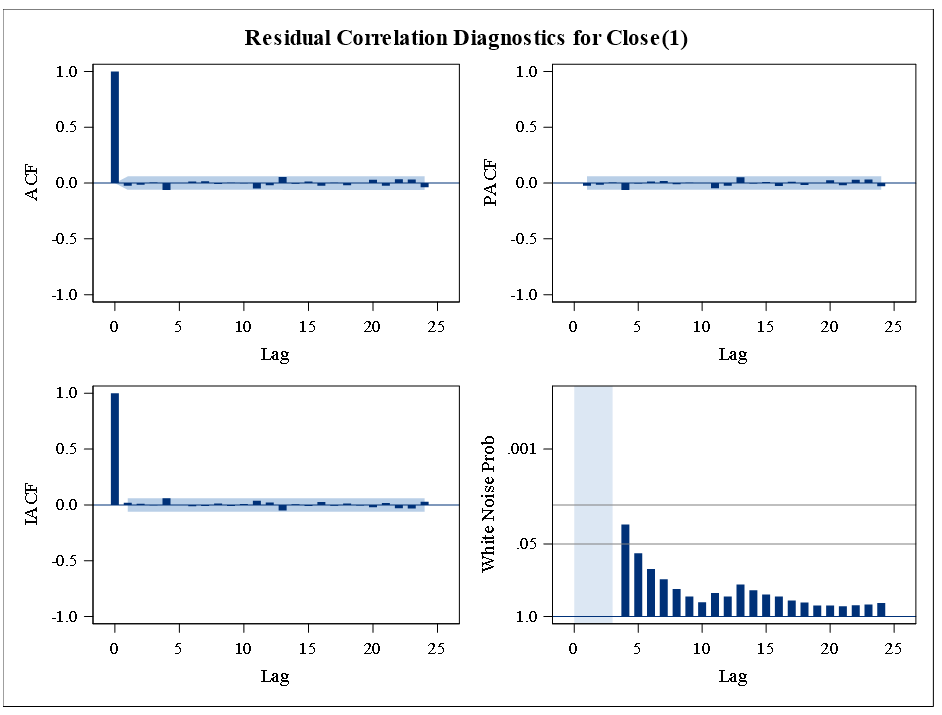
From the above residual plots we can see that there is very less significant correlation between the residuals across the first 24 lags and there is no white noise.

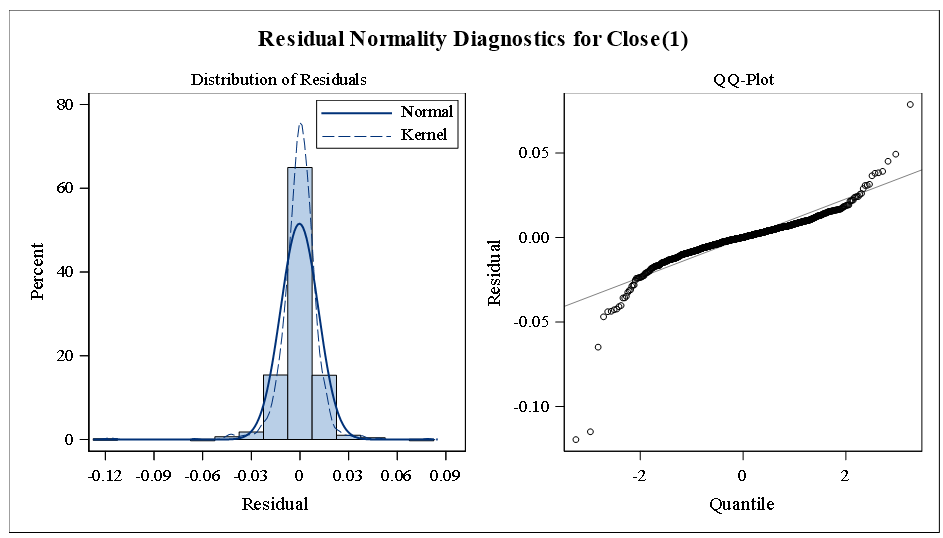
Now, we are modeling using the following parameters:

**Differencing = 1, Number of lags = AR (p)=2, MA (q)=1**

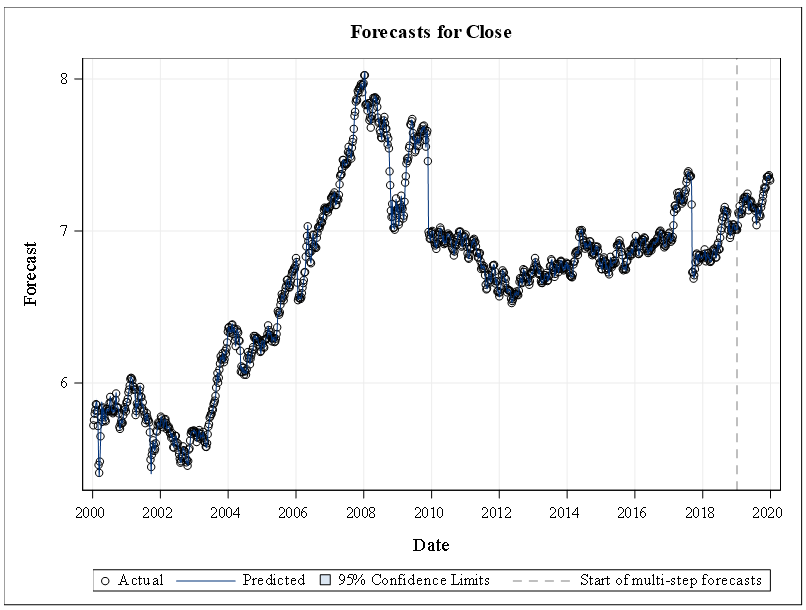
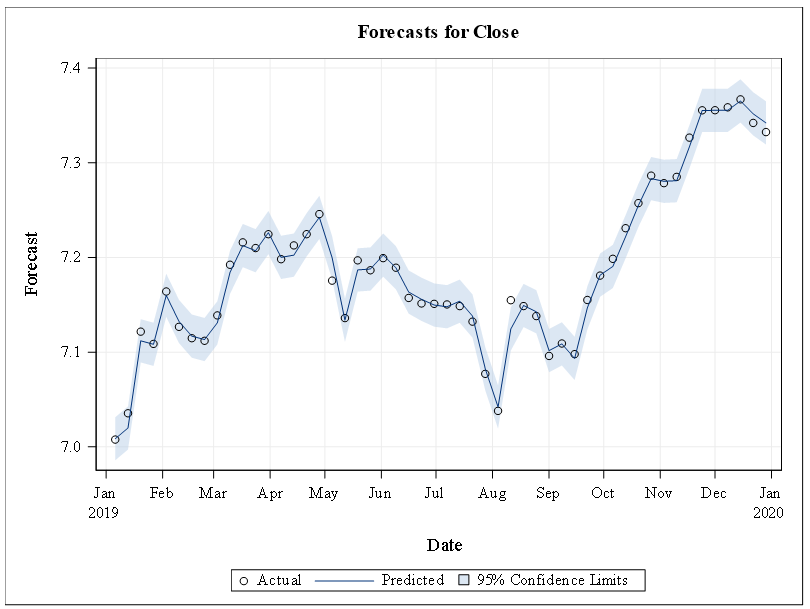
**Additional variables = Previous Close**

****



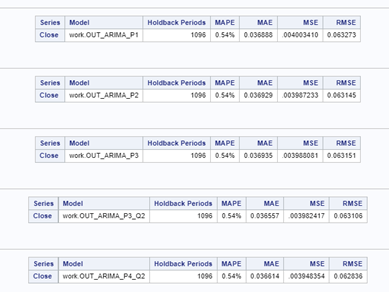


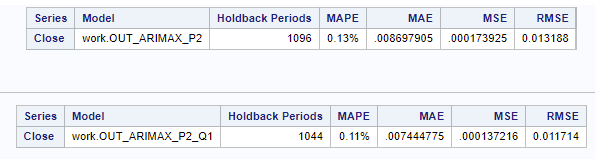
As the residuals are normally distributed we can say that our model fits well.



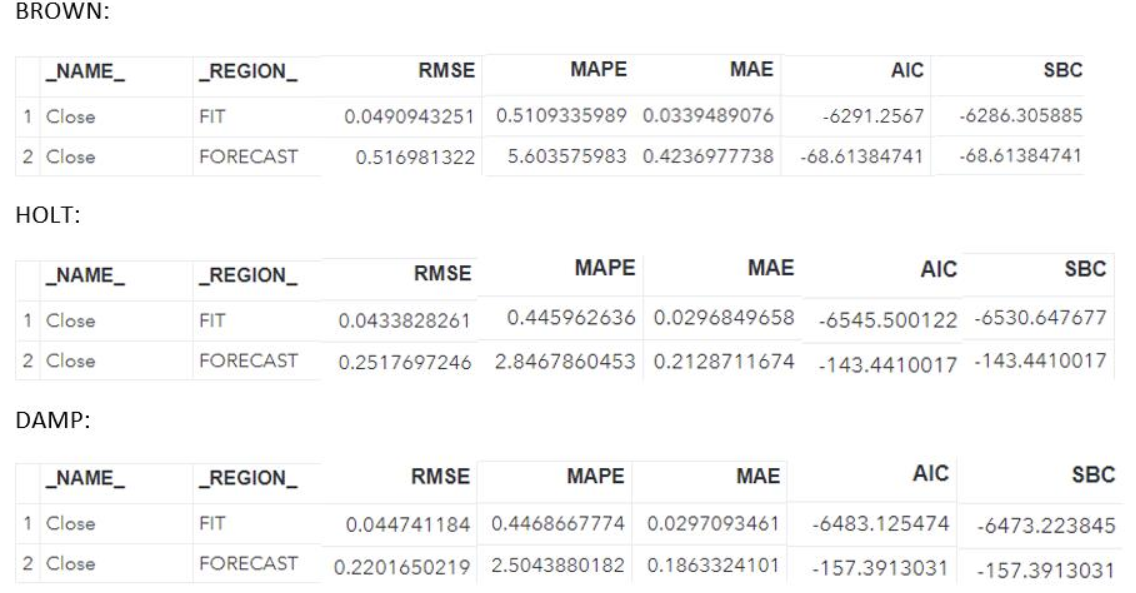
From the above forecast plot we can see that the actual and the predicted values are very close and the predicted value lies within the 95% confidence interval. This shows that our model is accurate with an MAPE value of 0.11%.

#### Model Comparison

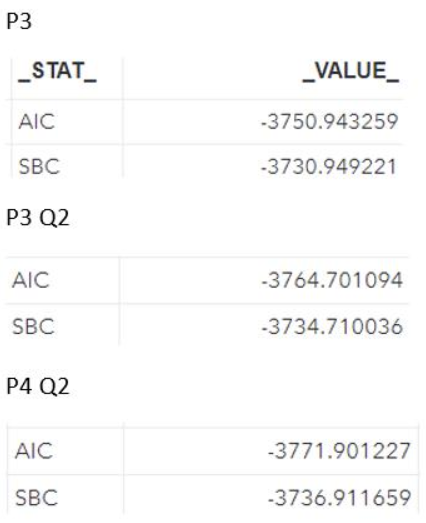
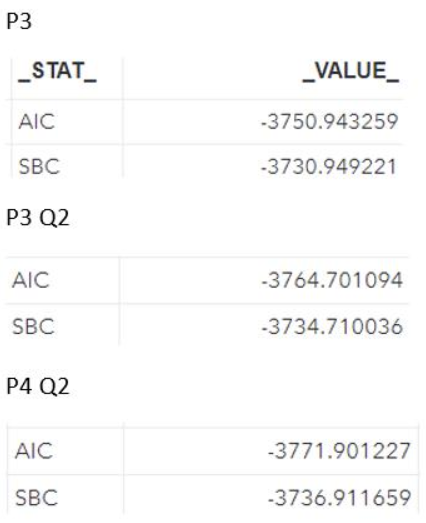
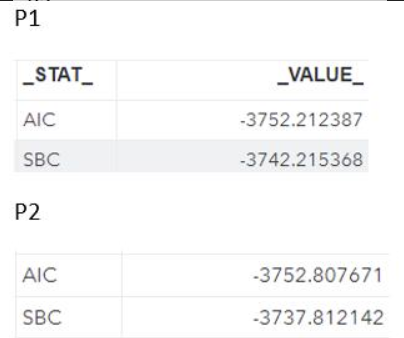




By seeing the above figure and the values of MAPE, MSE and RMSE we can say that ARIMA\_P4\_Q2 is performing better in the ARIMA models and ARIMAX\_P2\_Q1 is performing better in the ARIMAX models. Even though each ARIMA model is exhibiting the same accuracy, the MSE and RMSE values of ARIMA\_P4\_Q2 are lower. The MAPE, MSE and RMSE values for ARIMAX\_P2\_Q1 are lower when compared to other models making it the better model for forecasting the stock values.



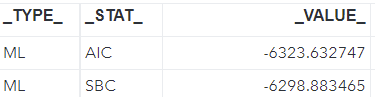
In the Exponential Smoothing model HOLT is performing better than the other two i.e; Double Brown and DAMP as the MAPE is slightly better than the others and the values of AIC and SBC are lower when compared to others. But we are not considering this model as the forecast produced by this model is not as effective as ARIMAX. Even though the forecast line and the actual data are in the 95% confidence level both of them are not moving closely which says that the error in predicting the values is a little higher than the other models such as ARIMAX.



Arimax P2



Arimax P2\_Q1



After considering the AIC and SBC values for the ARIMAX and ARIMA models which we’ve done we can double check that ARIMAX\_P2\_Q1 is performing better than others with the lowest AIC and SBC values of -6323.632747 and -6298.883465 which is second lowest. Even though the AIC and SBC values of ARIMAX\_P2 are lower, we have the values of MAPE, MSE, RMSE which are higher for this model and also there is no visible white noise. On the other hand ARIMAX\_P2\_Q1 has better MAPE and there is white noise and also lower residual error when compared to others. Thus saying that the ARIMAX\_P2\_Q1 is the best model for forecasting the stock values of Reliance Industries.

### Conclusions

Modeling stock prices is difficult. The ability to predict a stock’s future price, even one day ahead, would be valuable. Our hypothesis was that a reasonable model of the NIL stock would hold up through an event like the COVID-19 pandemic. The model that we created has done a good job in forecasting the values for 2020. The predicted values and the actual values are in the 95% confidence interval and are moving closely in the plot, which suggests that our model did a good job in predicting the values during COVID period. This shows that the stock forecasting by the model is effective and that the COVID-19 event did not disrupt our forecast of Reliance stocks.

### Recommendations

* The risk averse should invest in commodities. Reliance had a short term dip in stock price in March 2020, but quickly rebounded and has seen an upward trend since
* Use ARIMA P2\_Q1 time series models when forecasting and modeling Reliance during a pandemic
* Reliance could use this modeling information to prepare for a stock buyback during the forecasted dip where the stock price would be undervalued, resulting in raising stock prices and equity to shareholders which would increase profitability in the long term
* Financial advisors who develop ETF (electronically traded funds) based on the NIFTY 50 could use this information to increase the derivative indexes asset mixture of Reliance in lue of another pandemic, resulting in high profits for investors

### Citations

*Fortune 500 India: Largest company in India: Growth is life*. Reliance Industries Limited. (n.d.). <https://www.ril.com/OurCompany/About.aspx>.

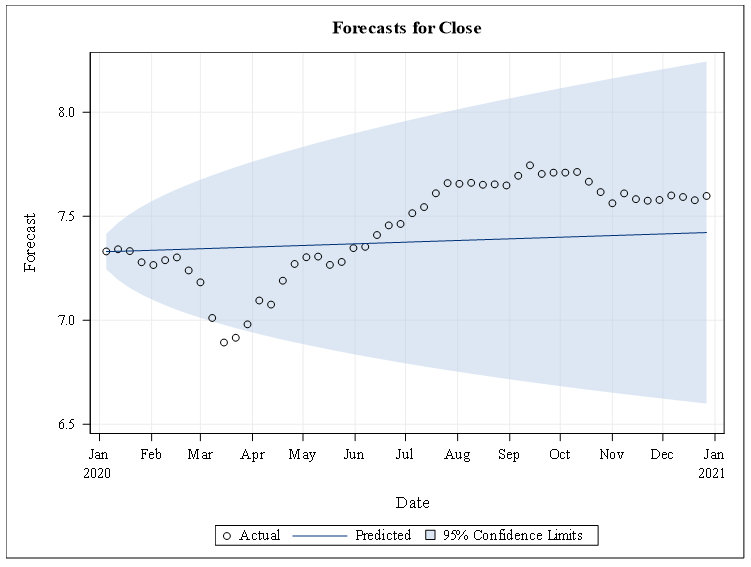
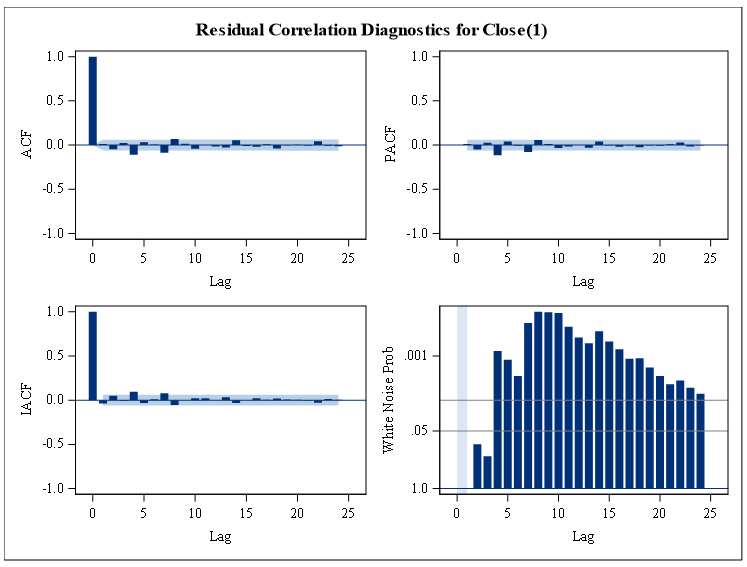
Reliance Industries Limited. (2021, May 1). <https://www.ril.com/ResponsetoCOVID-19.aspx>.

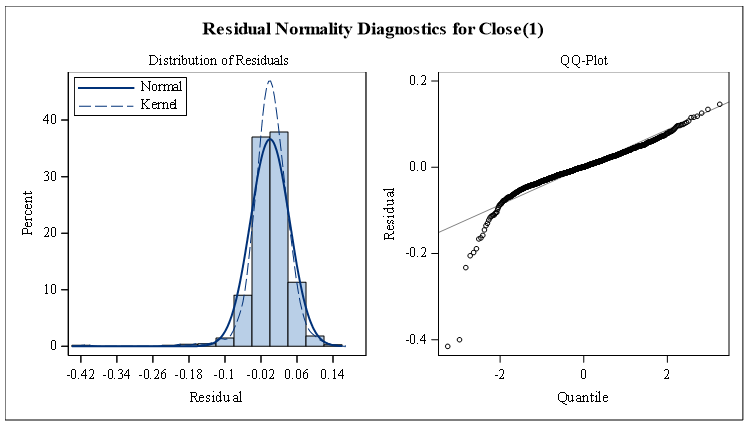
Vopani. (2021, May 1). *NIFTY-50 stock market Data (2000 - 2021)*. Kaggle. <https://www.kaggle.com/rohanrao/nifty50-stock-market-data>.

Appendix:

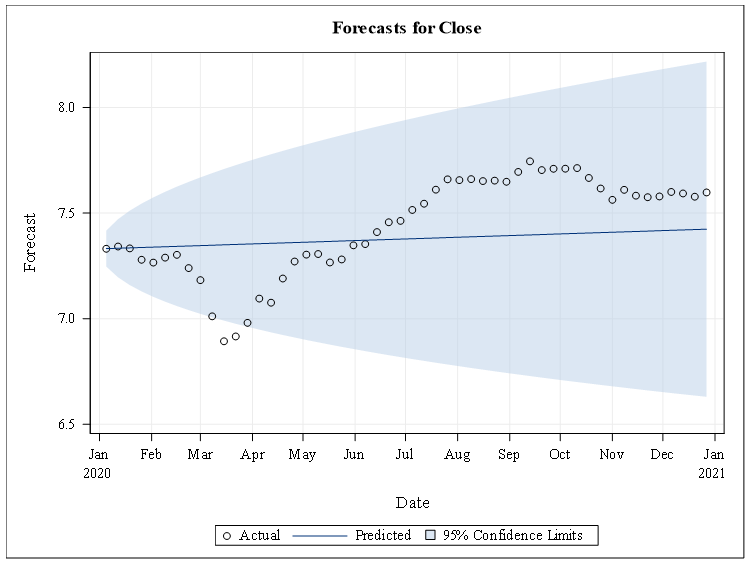
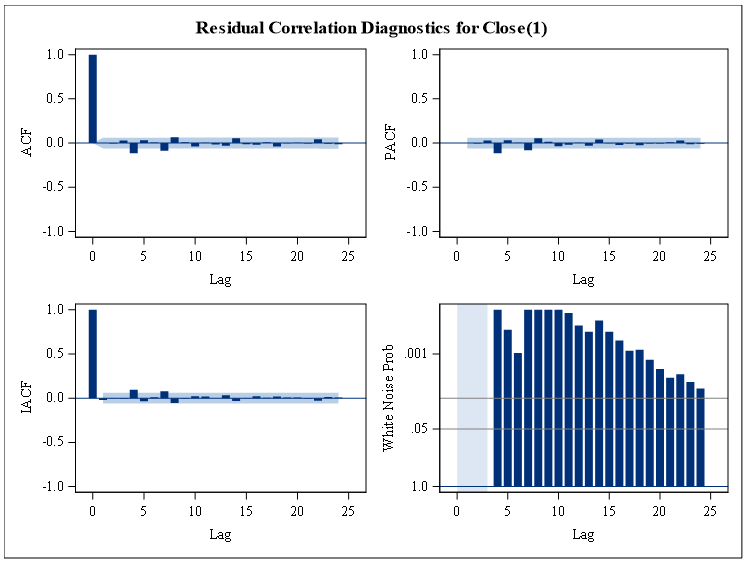
Other models we performed on our data:

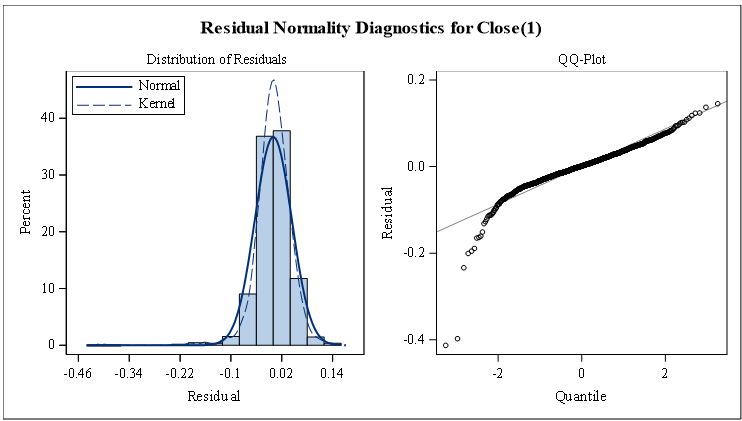
ARIMA - **Differencing = 1, Number of lags = AR = p= 2**

****

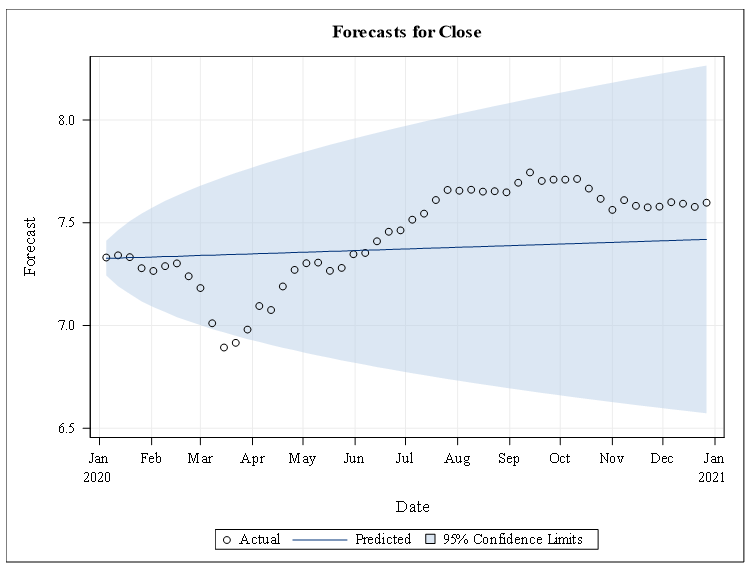
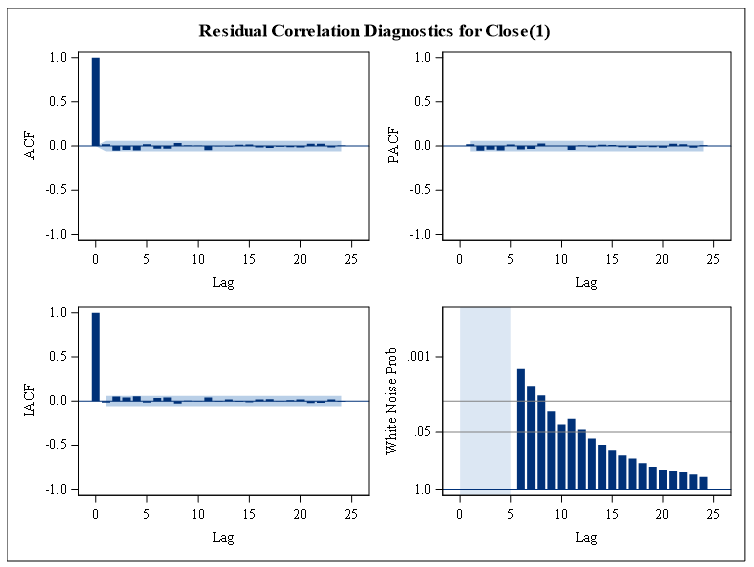
****

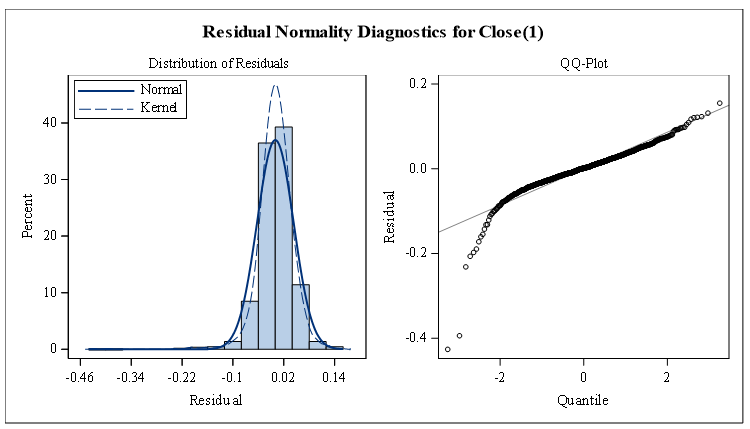
ARIMA - **Differencing = 1, Number of lags = AR = p= 3**

****

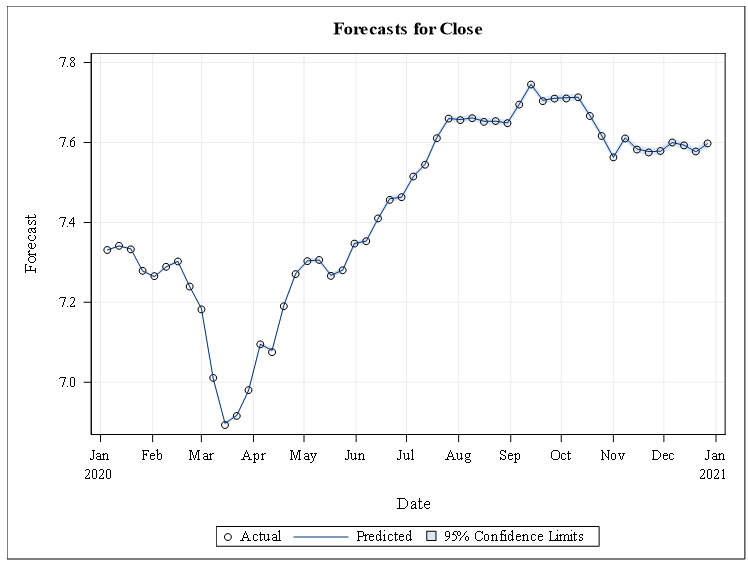
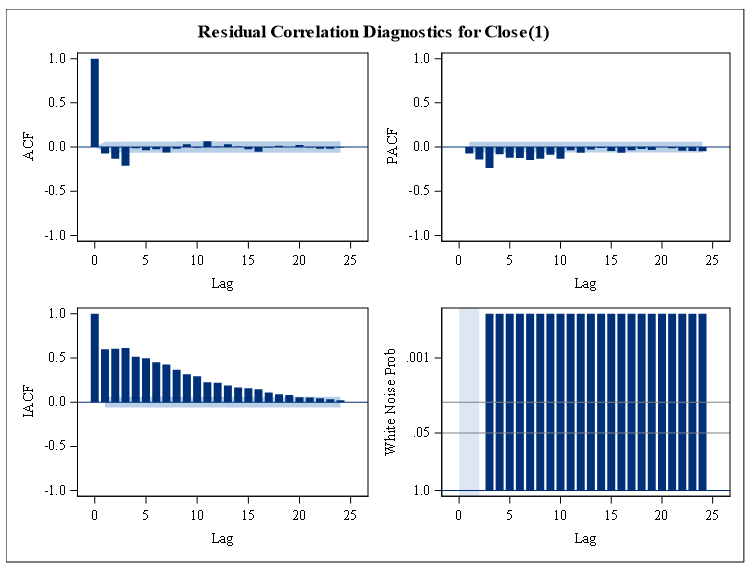
****

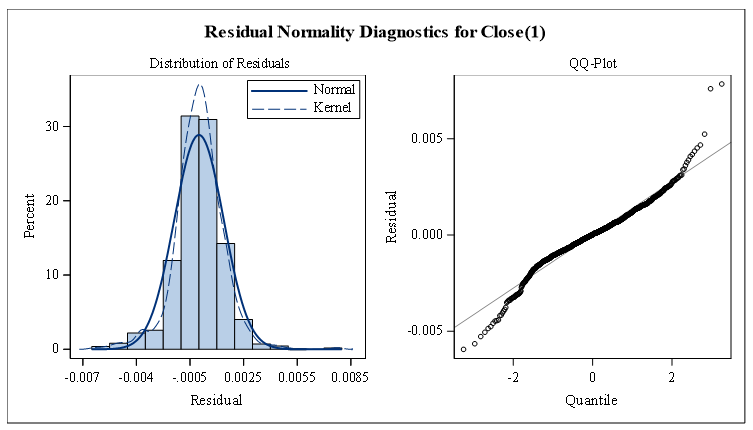
ARIMA - **Differencing = 1, Number of lags = AR = p= 3, q=2**

****

****

ARIMAX - **Differencing = 1, Number of lags = AR = p= 2 All Variables.**

****

****