PREDICTION OF STOCK PRICES USING TIME SERIES ANALYSIS

By Group 4

Animesh Gupta	2019B3AA0588H
Harsh Vardhan Gupta	2019B3A70630H
Aryan Kapadia	2019B3A70412H
Hitesh Garg	2019B3A70466H
Sujay Nigam	2019B3AA1267H
Dhruv Gupta	2019B3A70487H
Anand	2019B3A70613H
Chinmay Goyal	2019B3AA1290H

MAY 3, 2022

ECON F342: APPLIED ECONOMETRICS

BITS Pilani, Hyderabad Campus

Table of Contents

Section I: Introduction

Since their establishment, financial markets have been a hot topic of discussion amongst individuals. The interest has picked up even more during the pandemic. Financial markets across the world have seen heightened retail investor activity in the past 2 years. India has witnessed an explosion in the number of Demat accounts opened in the past few quarters. Switching to work from home enabled ample time for individuals, young and old, to dabble into the markets and make a quick buck. Thus, prediction of future stock prices is crucial for investors.

Advancements in technology have given rise to algorithmic trading which determines when to buy or sell a security using pre-programmed instructions based on variables like price, volume, timing etc. Both Investment banks and listed companies now rely heavily on algorithmic trading to establish how to act in financial markets. Interactions in financial markets becoming increasingly dependent on algorithms, in a sense, makes it easier for forecasting based on closing prices as the "human" element is decreasing. (Verheggen, 2017)

Researchers have employed various approaches to forecast future stock prices. ARIMA based time series models, Machine Learning models like K-Nearest Neighbours (KNN), Random Forest Model and Deep Learning models including RNN and LSTM (long-and short-term memory) have been used extensively to build models which forecast future stock prices.

This research paper uses ARIMA based time series models. The organization of this paper is such that Section II is on a brief look at related work in this field. Section III discusses the data and methodology, and Section IV presents the results and in-depth discussion. Section V concludes the paper and offers directions for future research.

Section II: Literature Review

Over the past decade, the forecasting of stock prices has emerged as an important field of research.

Debadrita Das (2014) in her research on 'Forecasting of Indian Stock Market using Time-series ARIMA Model' attempted to predict the future unobserved values of Indian Stock using closing prices of BSE Sensex for six years from 2006 to 2012, which have a very strong impact on the performance of the Indian economy. The study established the ARIMA model (1,0,1) as the most suitable model of BSE Sensex prediction and that the process is stationary and invertible. It also acknowledges the fact that case of any political turbulence or sudden change in Government fiscal, monetary or input policies will tend to create larger fluctuations in the Sensex price thereby deviating from the model since it does not capture the economic variables.

Sheik Mohammad Idrees et al. (2019) in their study on 'Prediction Approach for Stock Market Volatility on Time Series Data' used monthly closing prices for NIFTY Sensex from January 2010 to December 2016 to determine the best prediction model. They established ARIMA(0,1,0) with drift as the best model using ADF and L-Jung Box Test with roughly 5% deviation from the actual time series.

Along similar lines, **Prapanna Mondal et al. (2014)** studied the effectiveness of the ARIMA models in forecasting stock prices using the stock prices of 56 NSE listed companies, 8 companies from 7 sectors each for the time period between April 2012 to February 2014 and predicting next month's prices. It was determined that ARIMA(1,0,2) was the best fitting model with an accuracy of about 85%. ARIMA model was the most accurate in predicting the stock prices of the companies belonging to the FMCG sector, while the results for the banking and automobile sectors were relatively less accurate. The standard deviation was highest for the automobile, steel and banking sectors suggesting that the ARIMA model doesn't produce good results for stocks belonging to these sectors. For the IT sector, the standard deviation was average and an accuracy of about 90% was achieved using the forecasting model.

Ayodele et al. (2014) also conducted a stock price prediction study for two companies: Nokia Stock Index and Zenith Bank Stock: ARIMA(2,1,0) and ARIMA(1,0,1) were determined to be the best models respectively using the ADF test. The performance of both the models was quite impressive with very high accuracy and low standard deviation.

Jeffrey E Jarrett and Eric Kyper (2011) tried to model and analyse Chinese stock prices using the ARIMA analysis and show the impact of sudden economic interruptions. The study established that the financial crisis of 2008 had a significant setback on the Chinese stocks and manufacturing industry. It was also observed that Chinese equity securities had an autoregressive component which could be either short or long term(can be concluded only using long term financial modelling.

Section III: Data and Methodology

Description of Dataset

The analysis in this report used data sourced from Yahoo Finance. It consists of the daily historical values of the stock prices of Reliance Industries, with the following entries:

- Date: The date on which the observation was recorded.
- Open: The price at which the stock opened on that day in the stock market.
- High: The highest price of the stock on that day.
- Low: The lowest price of the stock on that day.
- Close: The price of the stock at the time of closing of the stock market for that day.
- Adjusted Close: The adjusted closing price amends a stock's closing price to reflect that stock's value after accounting for any corporate actions such as stock splits, dividend announcements, etc.
- Volume: The number of stocks traded on that day between daily open and close.

This report uses the values of Adjusted Closing Price to forecast the future stock prices using time series analysis.

Econometric Model

Since the data covers a period of 26 years (from 1996 to 2022), it is recommended that the seasonal component should be removed before proceeding with the time series regression. Additionally, the natural log of the adjusted closing price was taken since the original values suggested a multiplicative model. By using the log values, it becomes an additive series. Figure 1 shows the plot of the data before removal of the seasonal component.

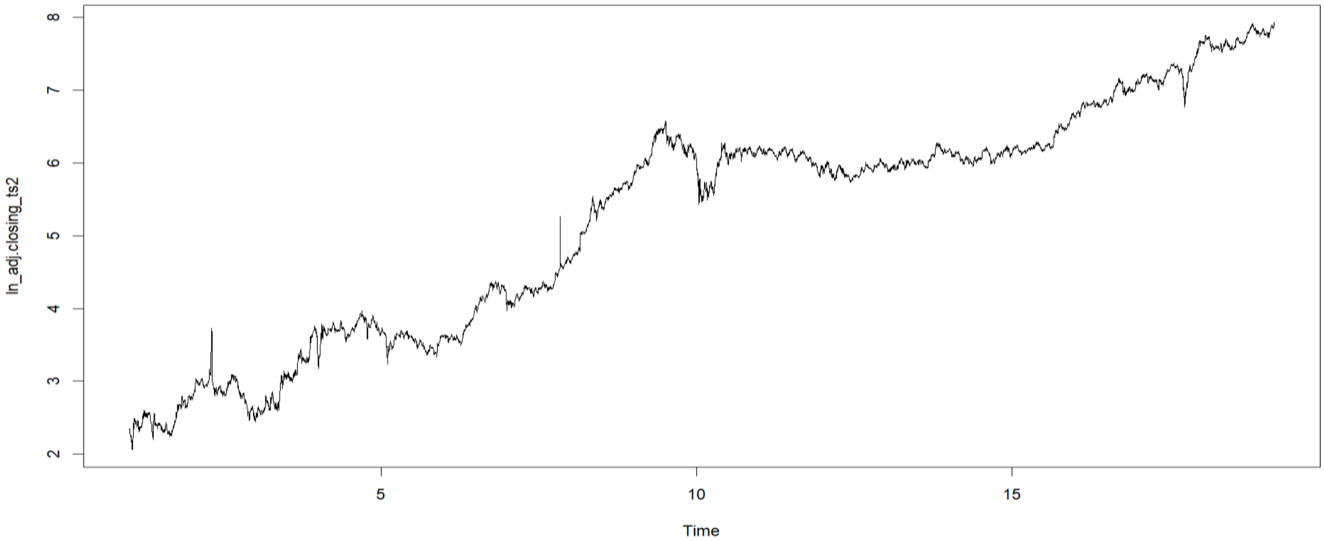

Figure 1: Plot of log of adjusted closing price over time

The decompose( ) function in R was used to find the different components of times series data i.e., seasonal, trend, cyclical, and random component. These are shown in Figure 2.

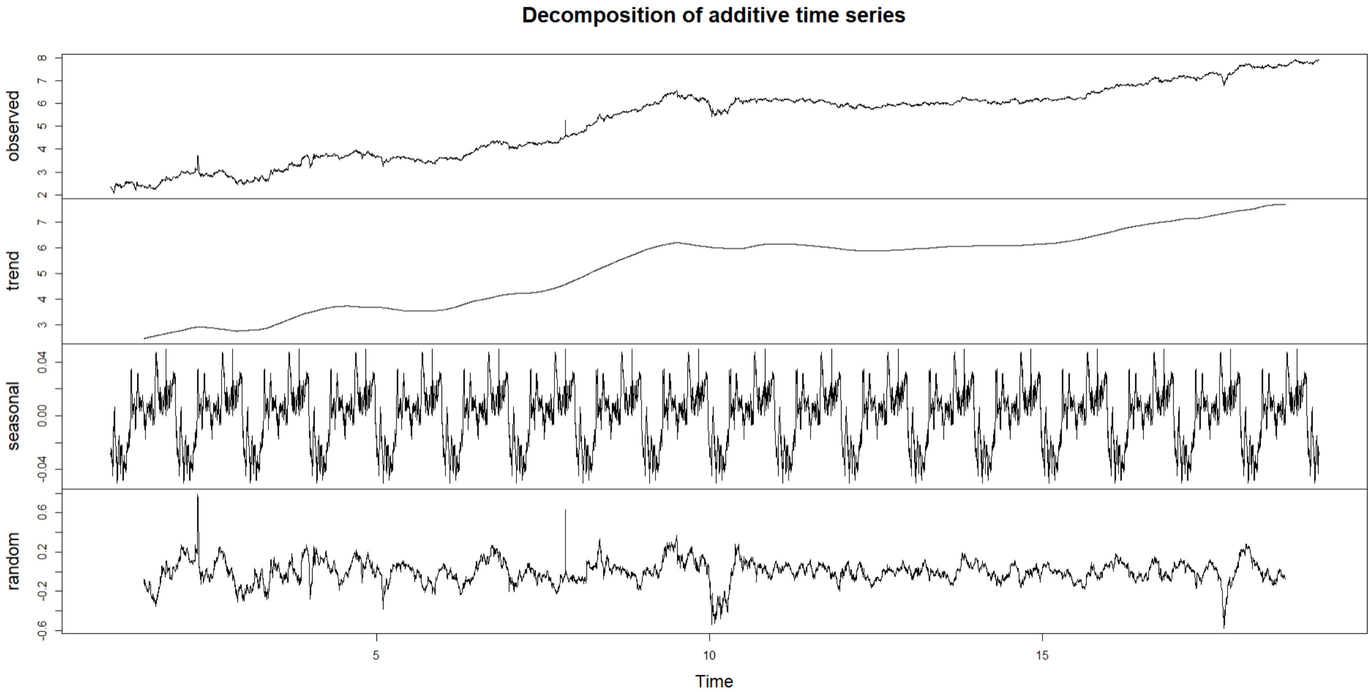

Figure 2: Different components of the time series data

The seasonal component was removed, and the resulting series is plotted in Figure 3.

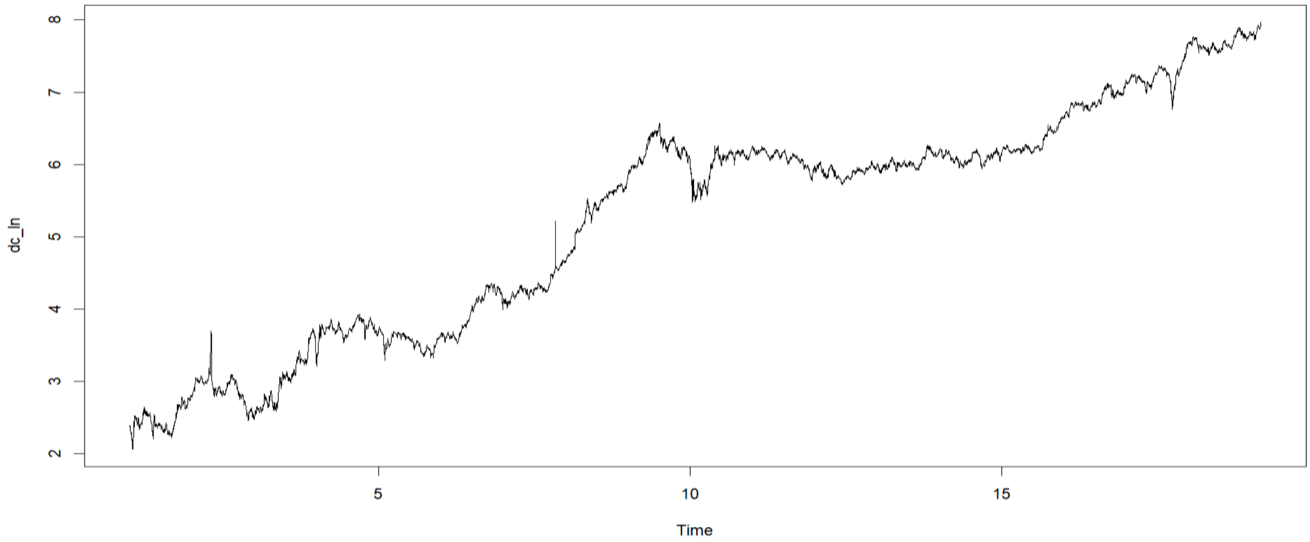

Figure 3: Deseasonalized time series

Finally, the auto.arima( ) function in R was used on this data. It suggested an ARIMA(0,1,1) model with drift, as shown in Figure 4.

```
> auto.arima(dc_ln)
Series: dc_ln
ARIMA(0,1,1) with drift

Coefficients:
          ma1   drift
      -0.0648   8e-04
s.e.   0.0123   3e-04

sigma^2 = 0.0007518:  log likelihood = 14431.93
AIC=-28857.87   AICc=-28857.86   BIC=-28837.47
```

Figure 4: Results of auto.arima( ) function

Thus, the econometric model for the analysis is:

$$\mathbf{Y_{adj.\ closing\ price\ at\ t} \sim ARIMA(0,1,1)}$$

The current series is not stationary because it is integrated of order 1. So, it needs to be differenced once, and the model becomes:

$$\mathbf{\Delta Y_{adj.\ closing\ price\ at\ t} \sim ARIMA(0,0,1)}$$

Now it is integrated of order 0 and hence it is stationary. Thus $\Delta Y_{\text{adj. closing price at t}}$ is an MA(1) process.

Section IV: Results and Discussion

Summary Statistics of the Model

```
> summary(dc_ln)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
  2.058   3.686   5.921   5.226   6.209   7.971 
```

Figure 5: Summary statistics of the variable

The minimum value of the deseasonalized logarithmic stock price of RELIANCE.NS is 2.058, while the maximum is 7.971. The first or lower quartile lies at 3.686 while the third or the upper quartile lies at 6.209. The mean and median lie at 5.226 and 5.921 respectively.

Checking for stationarity in the series

Since the data was integrated of order 1, it was differenced once and stored in the variable dc_ln_d1. The R function stationary.test( ) was used on this variable, which performs the Augmented Dickey Fuller test on time series data. The null hypothesis of this test is non-stationarity. The results of this test are given in Figure 6.

```
Type 2: with drift no trend 
      lag   ADF p.value
 [1,]   0 -86.8    0.01
 [2,]   1 -59.7    0.01
 [3,]   2 -48.8    0.01
 [4,]   3 -42.1    0.01
 [5,]   4 -37.7    0.01
 [6,]   5 -34.2    0.01
 [7,]   6 -34.1    0.01
 [8,]   7 -31.8    0.01
 [9,]   8 -29.3    0.01
[10,]   9 -27.0    0.01
[11,]  10 -25.6    0.01
```

Figure 6: Results of ADF test

Thus, the null is rejected at all lags, and the series is stationary.

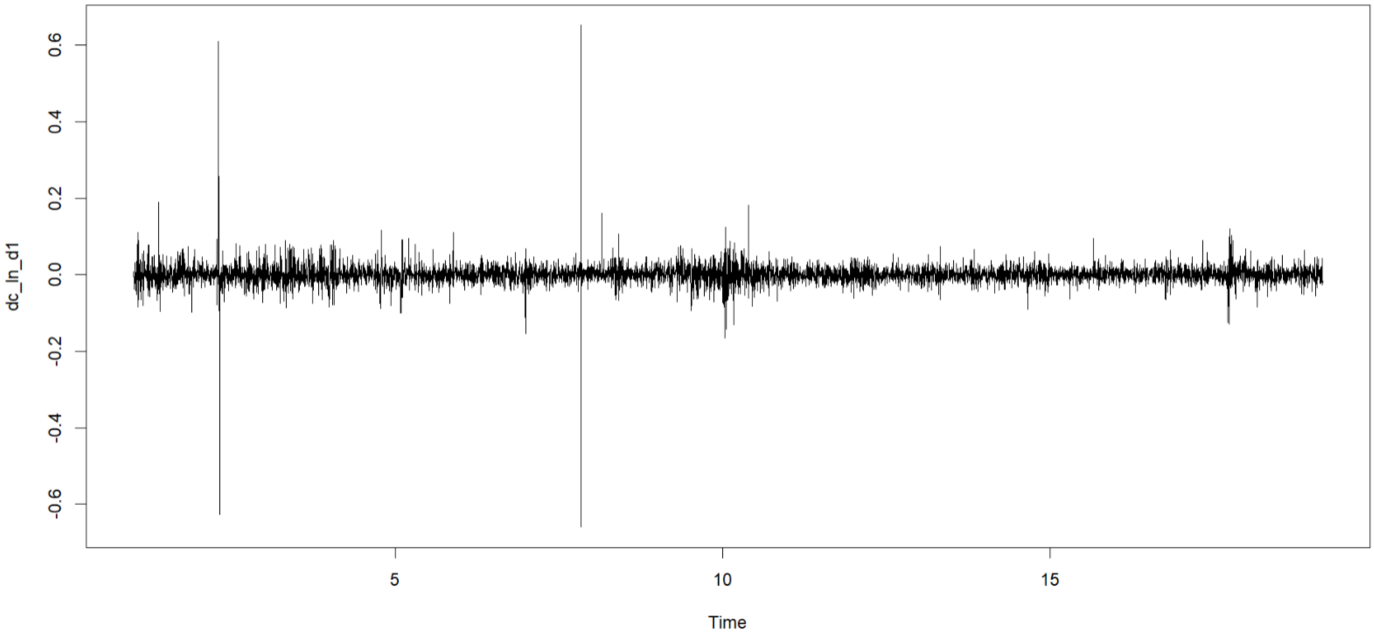

Figure 7: First differenced series is stationary

Model Parameters

```
> summary(reg_ma)
Series: dc_ln
ARIMA(0,1,1) with drift

Coefficients:
          ma1   drift
      -0.0648   8e-04
s.e.   0.0123   3e-04

sigma^2 = 0.0007518:  log likelihood = 14431.93
AIC=-28857.87   AICc=-28857.86   BIC=-28837.47

Training set error measures:
                       ME       RMSE        MAE          MPE      MAPE       MASE
Training set 5.505109e-07 0.02741236 0.01636308 -0.002372537 0.3675572 0.03997334
                     ACF1
Training set 0.0002524578
```

Figure 8: Results of the regression

Interpretation:

- ma1: This is the coefficient of the first lag of error term. Its value is -0.0648 which means that adjusted closing price at time t is negatively related to the error term at time t-1. The t-statistic for the significance test is -5.268. Thus, it is significantly different from zero at less 1% level.

- drift: The drift signifies the value about which the series is stationary. For the model of estimated in this report, the value of the drift is 0.0008. This is also significant at 1% level.

The IC values are also very low. It was compared to the IC values of ARIMA(1,1,0) model for the same data, which is shown in Figure 9.

```
> summary(reg_arima3)
Series: ln_adj.closing_ts
ARIMA(1,1,0) with drift

Coefficients:
          ar1  drift
      -0.0698  8e-04
s.e.   0.0123  3e-04

sigma^2 = 0.0007919:  log likelihood = 14235.09
AIC=-28464.17   AICc=-28464.17   BIC=-28443.78

Training set error measures:
                        ME       RMSE       MAE          MPE      MAPE       MASE
Training set -8.433729e-06 0.02813398 0.0163084 -0.002564813 0.3678772 0.03979168
                     ACF1
Training set 2.122242e-06
```

Figure 9: Results of an ARIMA(1,1,0) model

As shown, the IC values are higher than in the chosen model, which indicates that an ARIMA(0,1,1) model is a better suited model.

Conclusion from the model

The series, after first differencing, follows a Moving Averages model of order 1. This suggests that the value of the adjusted closing price does not depend on its past values, but rather it depends on the present value of the error term and the first lag of the error term i.e., u_t and u_{t-1}.

Checking for Time Series assumptions

1. Stationarity of the series

As shown in Figure 7, the series becomes stationary after the first difference. The ADF test conducted on the first differenced series also confirms the stationarity (Figure 6).

2. No Serial Autocorrelation between the error terms of the forecast

This can be checked by plotting the correlogram of the residuals after forecasting future values using the model (Figure 10). As can be seen, the error residuals are not correlated with each other after the 0^{th} lag.

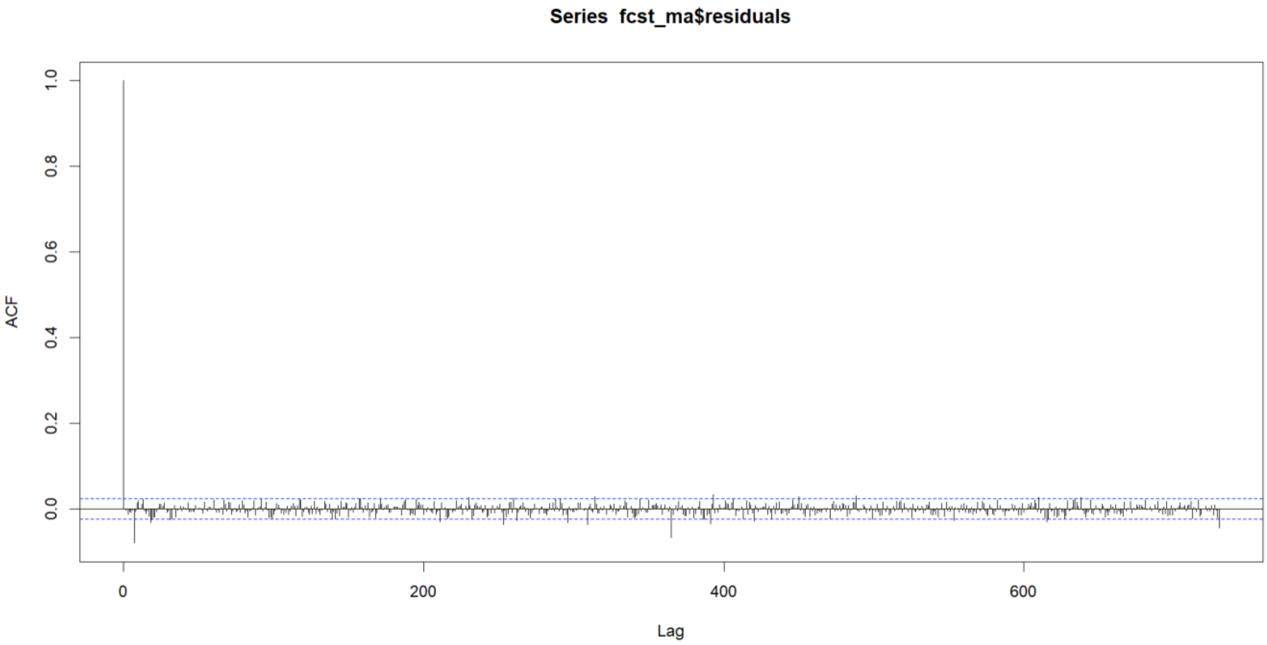

Figure 10: ACF of residuals

A formal test for checking this condition was also conducted which is the Ljung-Box test. The results for the same are shown in Figure 11.

```
        Box-Ljung test

data:  fcst_ma
X-squared = 2.1333, df = 1, p-value = 0.1441
```

Figure 11: Results of Ljung-Box test

The null hypothesis of Ljung-Box test is no serial autocorrelation between the residuals. The p-value of 0.1441 implies that the null hypothesis cannot be rejected. Thus, there is no serial autocorrelation between the residuals.

3. Normality of residuals

From Figure 12, it can be seen that the residuals form a normal shape though, there is high kurtosis.

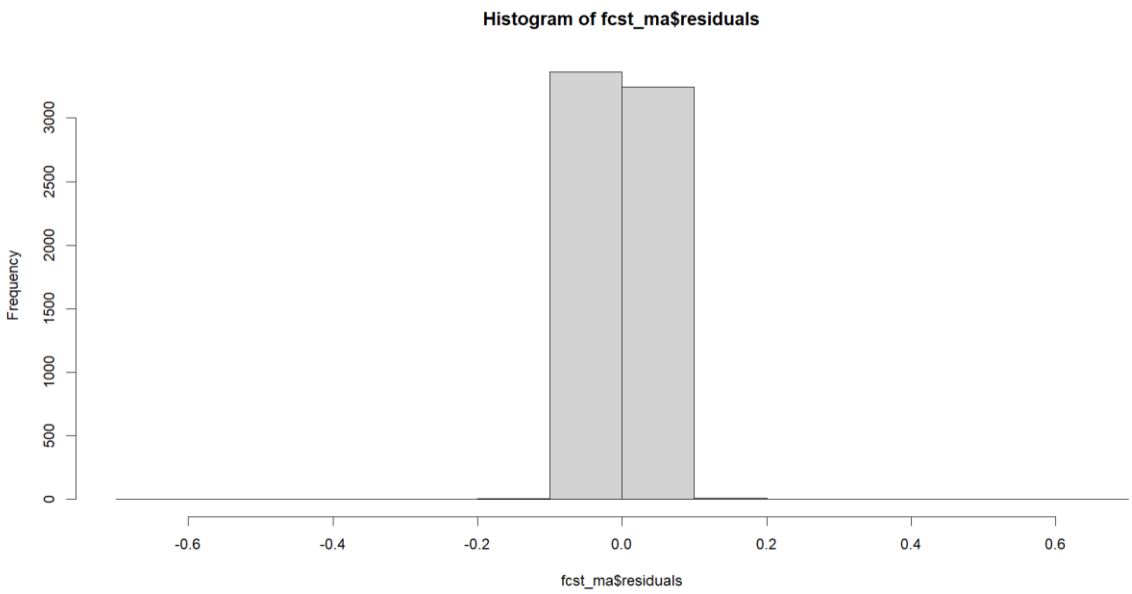

Figure 12: Histogram of Forecast Residuals

Thus, all the required assumptions are satisfied, and no remedial measures are required.

Forecasting Values Using the Model

Point forecasts of the natural log of the adjusted closing price for 5 days in the future were estimated using the forecast function in R. The values were within a Confidence Interval of 99.5%, and are shown in Figures 13, 14, and 15.

```
> forecast(reg_ma,h=5,level=99.5)
         Point Forecast Lo 99.5  Hi 99.5
19.15890       7.930798 7.853834 8.007763
19.16164       7.931636 7.826256 8.037015
19.16438       7.932473 7.804856 8.060090
19.16712       7.933310 7.786793 8.079827
19.16986       7.934148 7.770904 8.097391
```

Figure 13: Point Forecasts of next 5 days

Figure 14: Point Forecasts of next 5 days graphically

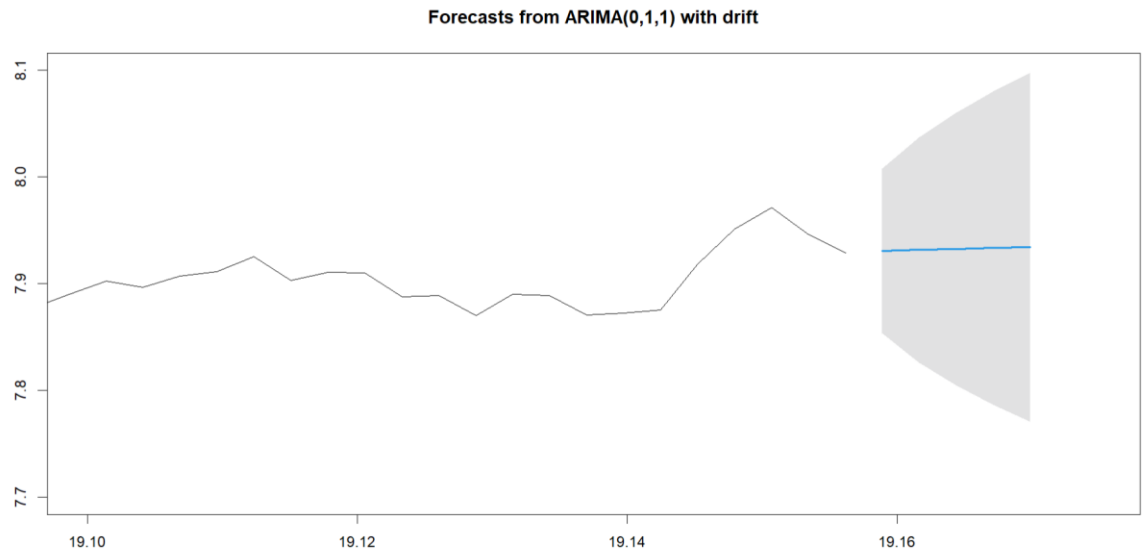

Figure 15: Point Forecasts of next 5 days graphically (zoomed in)

Interpretation of the forecasted values

The forecasted values are of the natural log of adjusted closing price. Thus, the actual predicted values are found by taking their exponential as given below:

```
     26-04-2022 27-04-2022 28-04-2022 29-04-2022 2-05-2022
[1,]   2781.647   2783.977   2786.309   2788.643  2790.979
```

Figure 16: Actual 5 day forecast of adjusted closing price

These forecasted values were tested against the actual values of Reliance Industries Adjusted Closing Prices for the next 5 days. The difference between the two are given in Figure 17. It is to be noted that the values for 30-04-2022 and 01-05-2022 were padded by the graphing software; there is no real data for those dates since the exchange remains closed on non-business days.

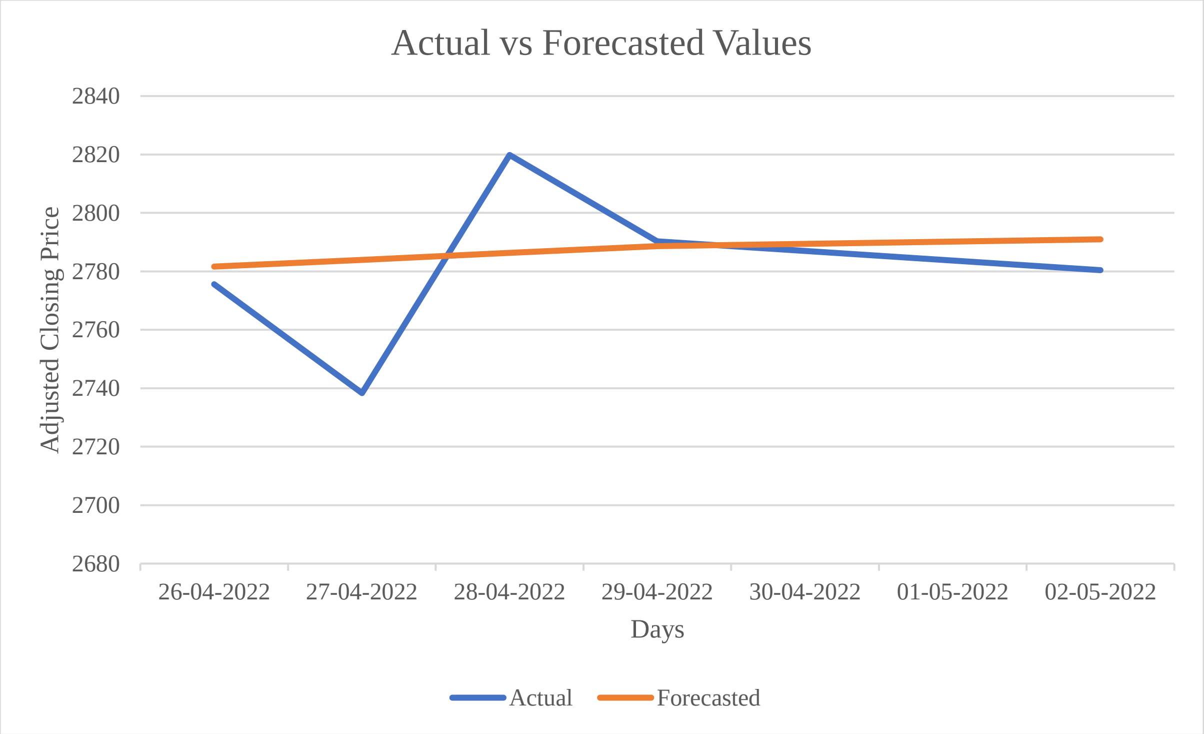

Figure 17: Actual vs Forecasted Values

The estimated model can forecast the values within an average relative error rate of 0.7%.

Section V: Conclusion

This report focused on predicting the stock prices of Reliance Industries Ltd using time series analysis. The selected variable for this analysis was the natural log of the adjusted closing price of the stock, observed daily for 26 years. An ARIMA(0,1,1) model was established as the best model for this purpose, after the series was deseasonalized, based on AIC and BIC values. This model satisfied the assumptions required for time series analysis: stationarity of the series (using ADF test), no serial autocorrelation (using correlogram and Ljung-Box test) of the forecast residuals, and normality of the residuals (using histogram of residuals).

The model was used to forecast values of the adjusted closing price for 5 business days after the last date in the training sample. These values were compared with the true values obtained from the same source (Yahoo Finance), and it was determined that the model had a relative error rate of 0.7%. Although the accuracy of the model is very high, the testing sample is relatively small, and hence this accuracy might not be persistent for testing samples of longer durations.

This model also does not consider the firm specific factors which can affect the stock price externally, like change in management, employee strikes, etc. Moreover, drastic changes in the government's fiscal and monetary policy are also not captured by the model. These can cause the actual values to deviate from the forecasted values more than expected, thus reducing the accuracy of the model.

References

1. Banerjee, Debadrita. (2014). Forecasting of Indian stock market using time-series ARIMA model. 2014 2nd International Conference on Business and Information Management, ICBIM 2014. 131-135. 10.1109/ICBIM.2014.6970973.

2. Idrees, Sheikh & Alam, Afshar & Agarwal, Parul. (2019). A Prediction Approach for Stock Market Volatility Based on Time Series Data. IEEE Access. PP. 1-1. 10.1109/ACCESS.2019.2895252.

3. Mondal, Prapanna & Shit, Labani & Goswami, Saptarsi. (2014). Study of Effectiveness of Time Series Modeling (Arima) in Forecasting Stock Prices. International Journal of Computer Science, Engineering and Applications. 4. 13-29. 10.5121/ijcsea.2014.4202.

4. Adebiyi, Ayodele & Adewumi, Aderemi & Ayo, Charles. (2014). Stock price prediction using the ARIMA model. Proceedings - UKSim-AMSS 16th International Conference on Computer Modelling and Simulation, UKSim 2014. 10.1109/UKSim.2014.67.

5. Kyper, Eric. (2011). ARIMA Modeling with Intervention to Forecast and Analyze Chinese Stock Prices. International Journal of Engineering Business Management. 3. 10.5772/50938.

6. Yahoo Finance. (1996–2022). *RELIANCE.NS Historical Data* [Dataset]. https://finance.yahoo.com/quote/RELIANCE.NS/history/

7. Verheggen, R. (2017, November). *The rise of Algorithmic Trading and its effects on Return Dispersion and Market Predictability*. http://arno.uvt.nl/show.cgi?fid=145161