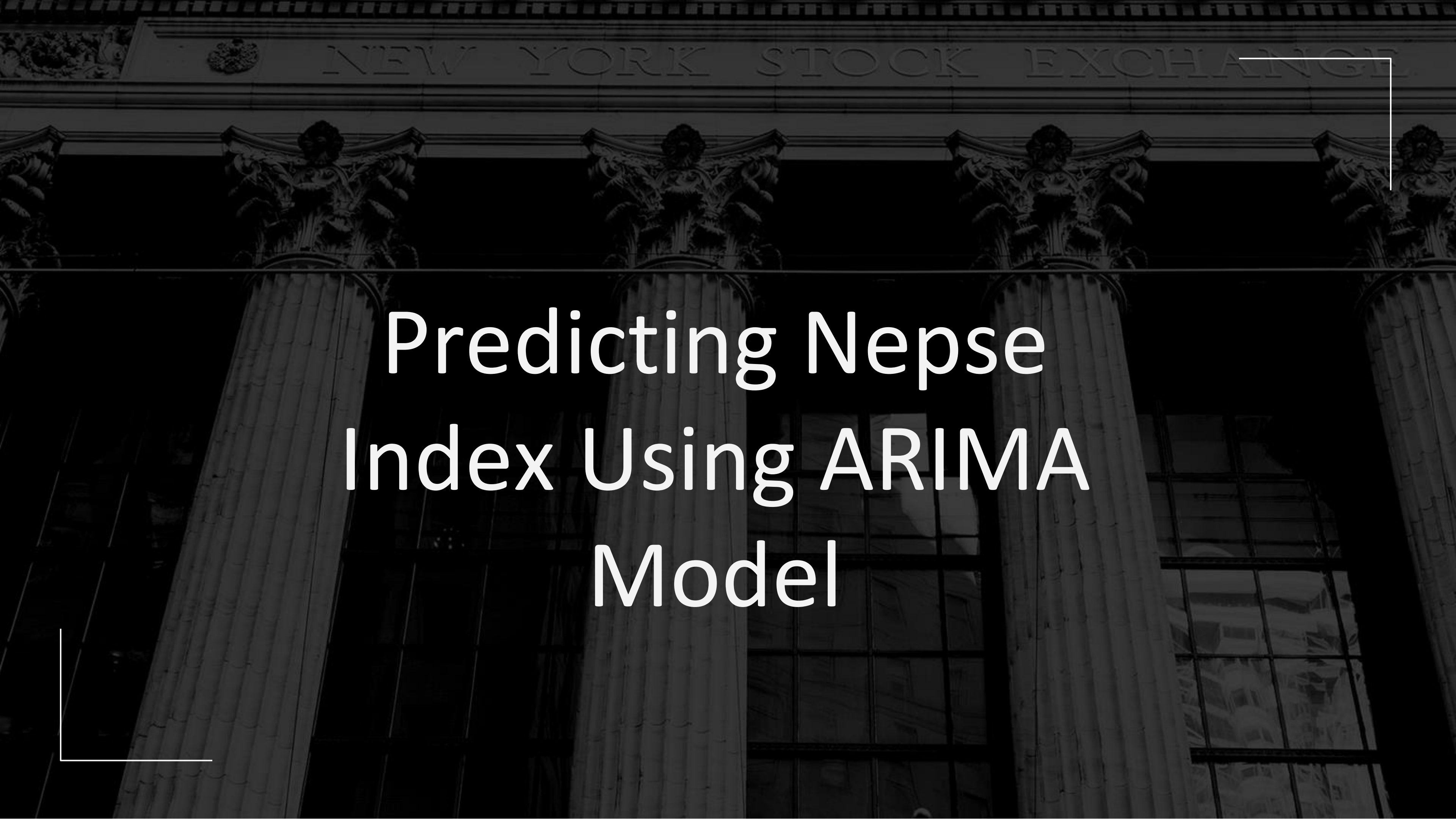


Predicting Nepse Index Using ARIMA Model

Presenter/Author

Avinash Maskey

MCIS - III

202802

Table of Contents

01 What is Time Series?

Introduction

- Time series is a sequence of observations recorded at regular time intervals. Depending on the frequency of observations, a time series may typically be hourly, daily, weekly, monthly, quarterly and annual.

- Sometimes, you might have seconds and minute-wise time series as well, like, number of clicks and user visits every minute etc.

- So, **what does analyzing a time series involve?** Time series analysis involves understanding various aspects about the inherent nature of the series so that you are better informed to create meaningful and accurate forecasts.

What is ARIMA MODEL?

02

Introduction

- AutoRegressive Integrated Moving Average or popularly known as ARIMA is a very widely used time series forecasting technique.
- Before starting prediction with ARIMA let us understand the concept of stationary. A time-series prediction is done only if the dataset is stationary.
- A dataset is said to be stationary if its mean and variance remains constant over time. A stationary dataset does not have trend or seasonality.
- When forecasting using time series models, we assume that each data point is independent of the other and this can be confirmed if the series is stationary.
- A basic non-seasonal ARIMA model is identified as an ARIMA (p, d, q) model.

03

ARIMA Algorithm

Understanding ARIMA Model

Let us look at the ARIMA model in detail. The ARIMA algorithm is made of the following components:

- The **AR** stands for Auto Regression which is denoted as p, the value of p determines how the data is regressed on its past values.

- The **I** stand for Integrated or the differencing component which is denoted as d, the value of d determines the degree of difference used to make the series stationary.

- The **MA** stands for Moving Average which is denoted as q, the values of q determine the outcome of the model depends linearly on the past observations and the same goes for the errors in forecasting as they also vary linearly.

- Mathematically,

$$Y_t = \alpha + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + .. + \beta_p Y_{t-p} \epsilon_t + \phi_1 \epsilon_{t-1} + \phi_2 \epsilon_{t-2} + .. + \phi_q \epsilon_{t-q}$$

Where,

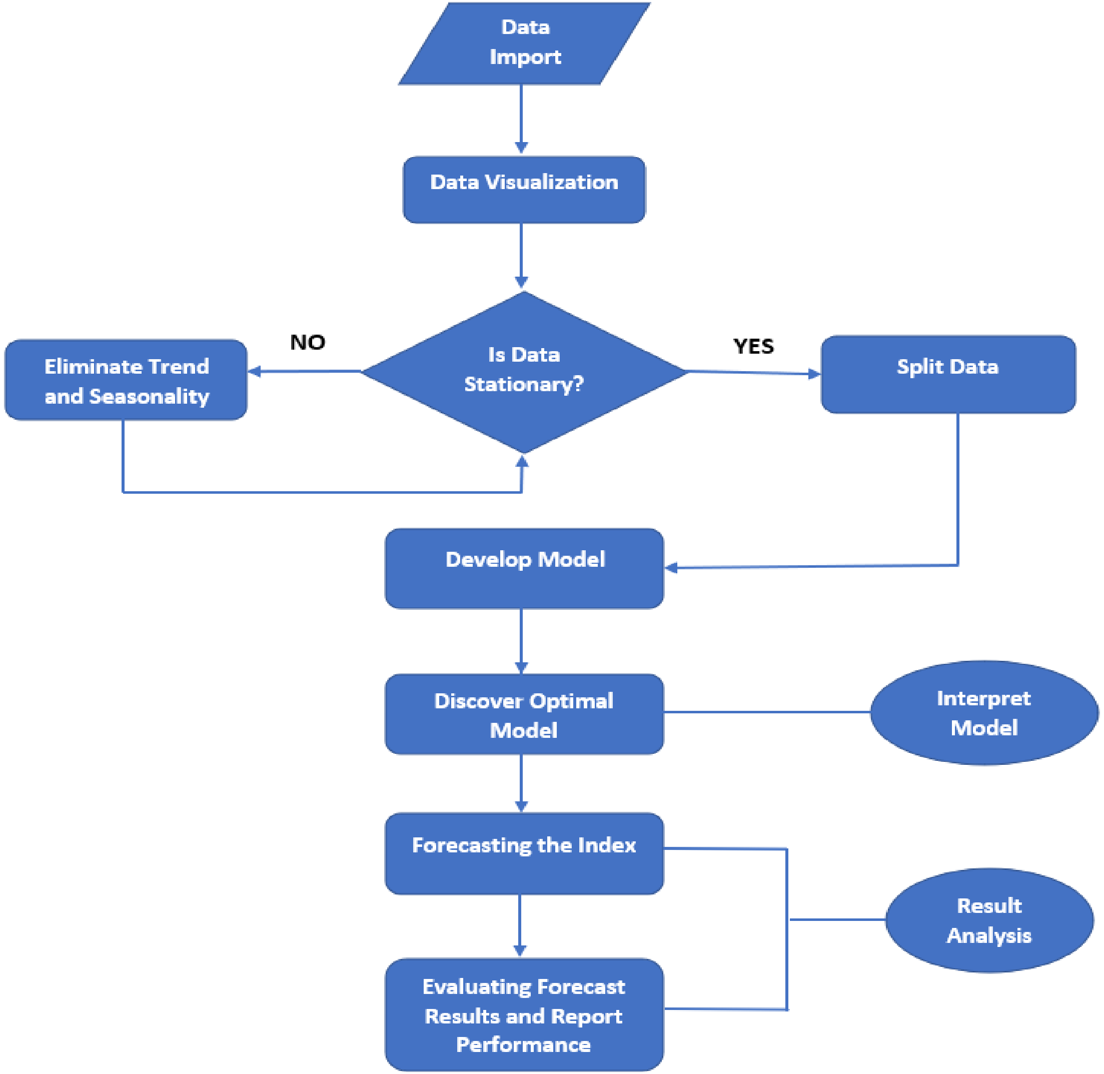
Predicted Y(t) = Constant + Linear combination Lags of Y (up to p lags) + Linear Combination of Lagged forecast errors (up to q lags)

04

Methodology

Methodology

- In order to achieve the set of objectives of the study an analytical research design has been adopted.
- The research has been designed in such a way that the collection, analysis and interpretation of the secondary data related to the study may be easier and efficient while drawing conclusions.

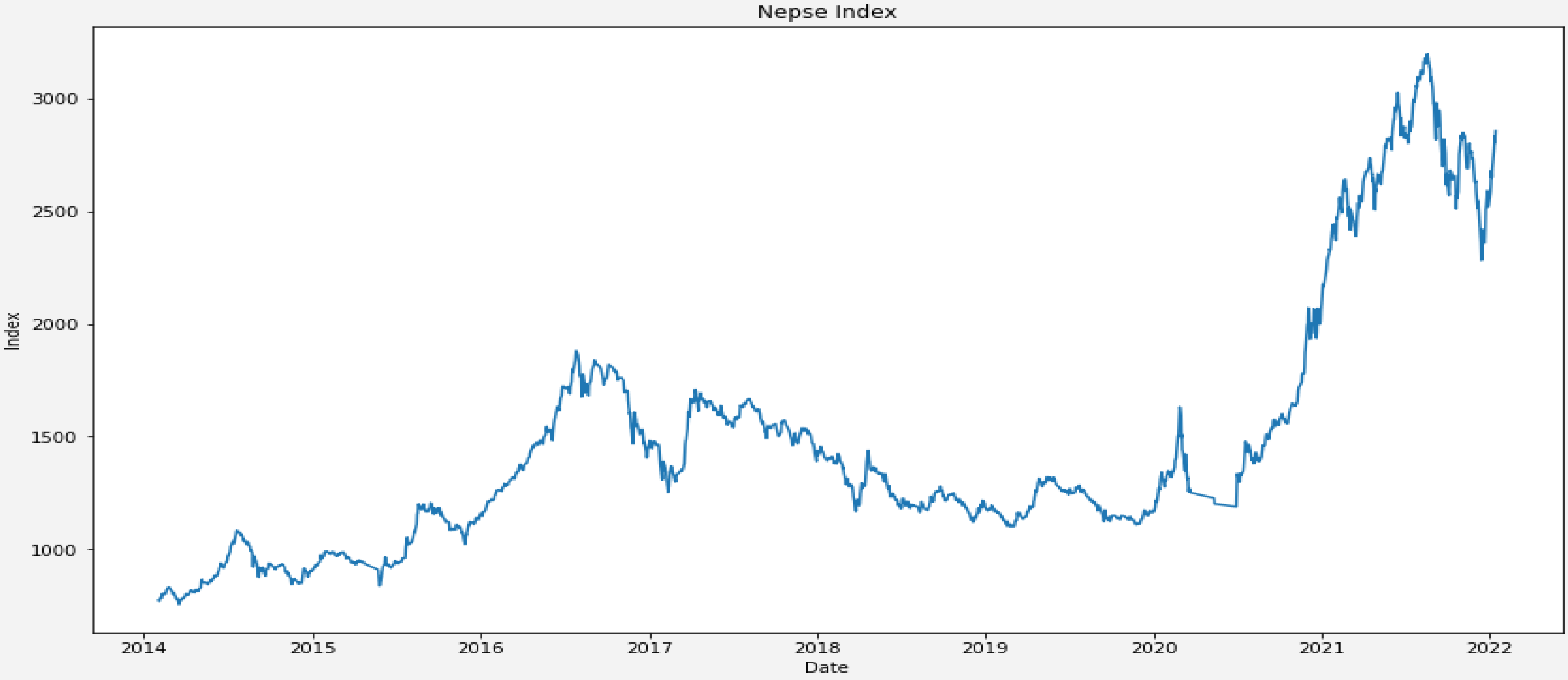


4.1 Data Import

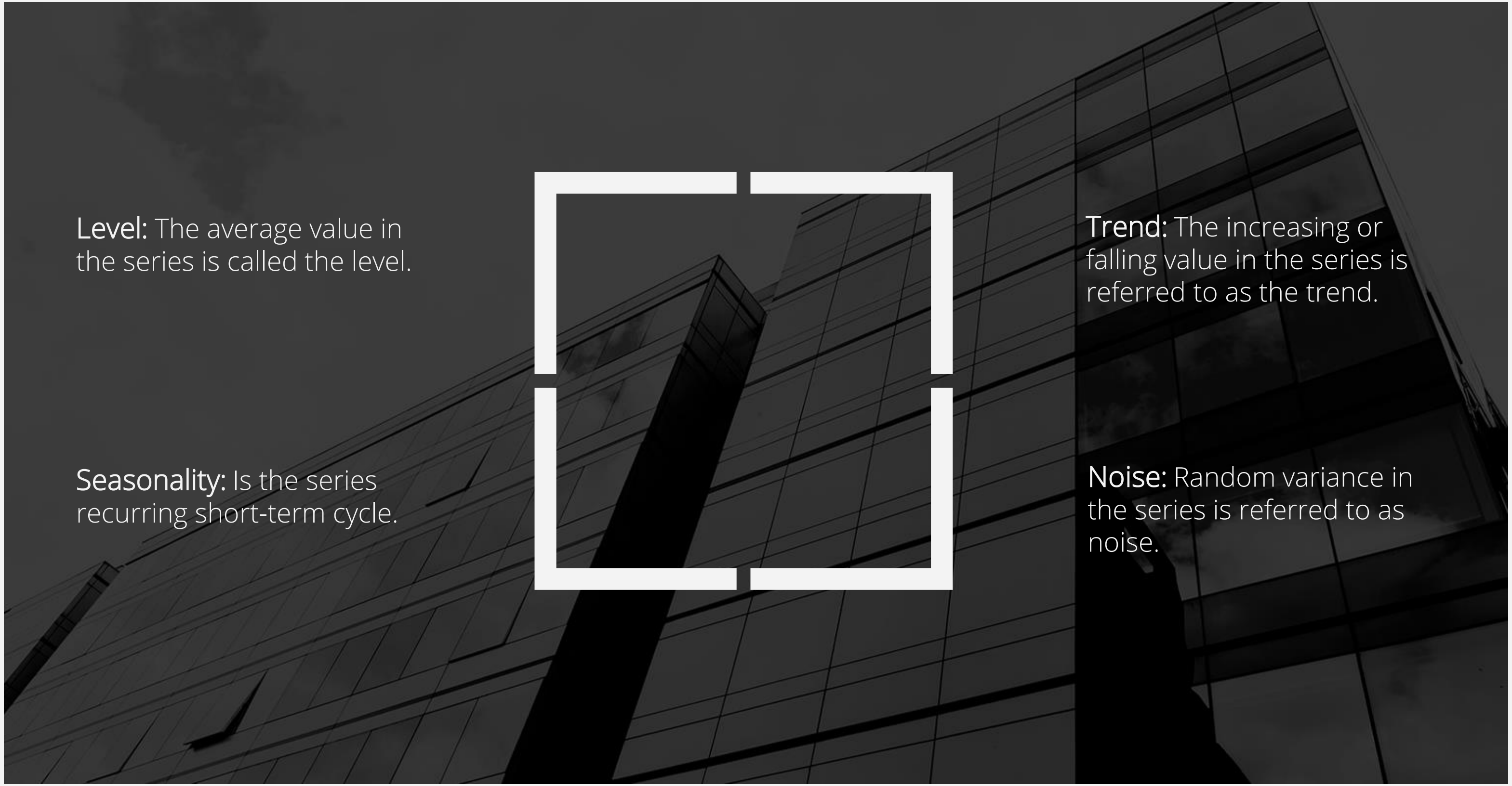
The data imported are Daily Closing Index of Nepal Stock Exchange (Nepse) and reported from 2nd February 2014 to 13th January 2022.

4.2 Data Visualization

Visualizing the stock's daily closing index.



Few Terminologies



Level: The average value in the series is called the level.

Trend: The increasing or falling value in the series is referred to as the trend.

Seasonality: Is the series recurring short-term cycle.

Noise: Random variance in the series is referred to as noise.

4.3 Stationarity Test

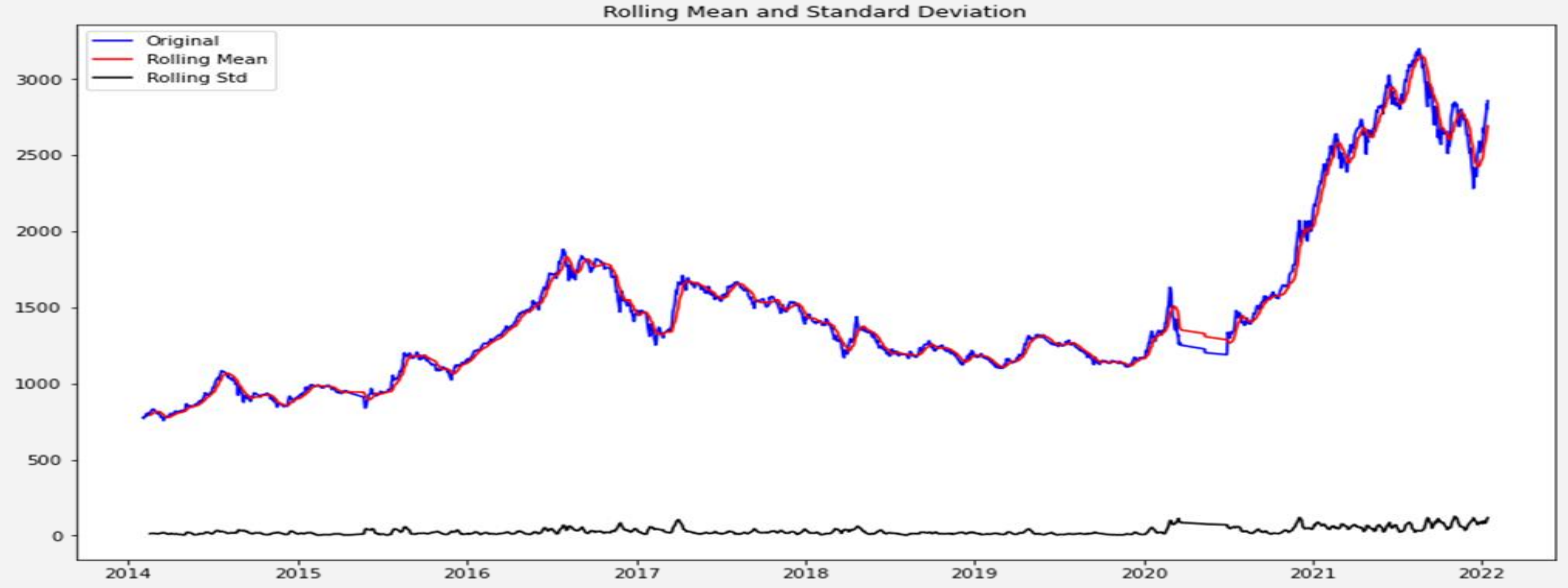
- Stationarity is a property of a time series. A stationary series is one where the values of the series is not a function of time. That is, the statistical properties of the series like mean, variance and autocorrelation are constant over time.

- One of the most widely used statistical tests is the Dickey-Fuller test.

- It can be used to determine whether or not a series has a unit root, and thus whether or not the series is stationary. This test's null and alternate hypotheses are:

    I. **Null Hypothesis:** The series has a unit root.
    II. **Alternate Hypothesis:** The series has no unit root.

- If the null hypothesis is not rejected, the series is said to be non-stationary.



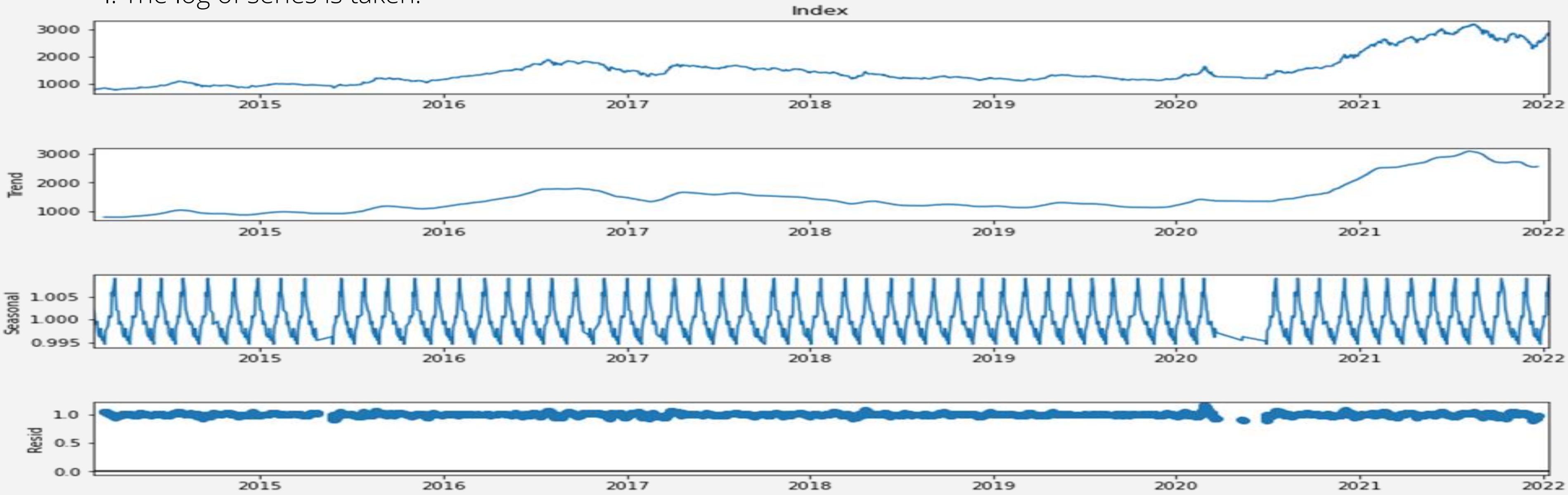
Results of Dickey Fuller Test:

Test Statistics	0.085856
p-value	0.965082
No. of lags used	19.000000
Number of observations used	1778.000000
critical value (1%)	-3.434033
critical value (5%)	-2.863167
critical value (10%)	-2.567636

4.3.1 Stationarity Test - Eliminate Trend

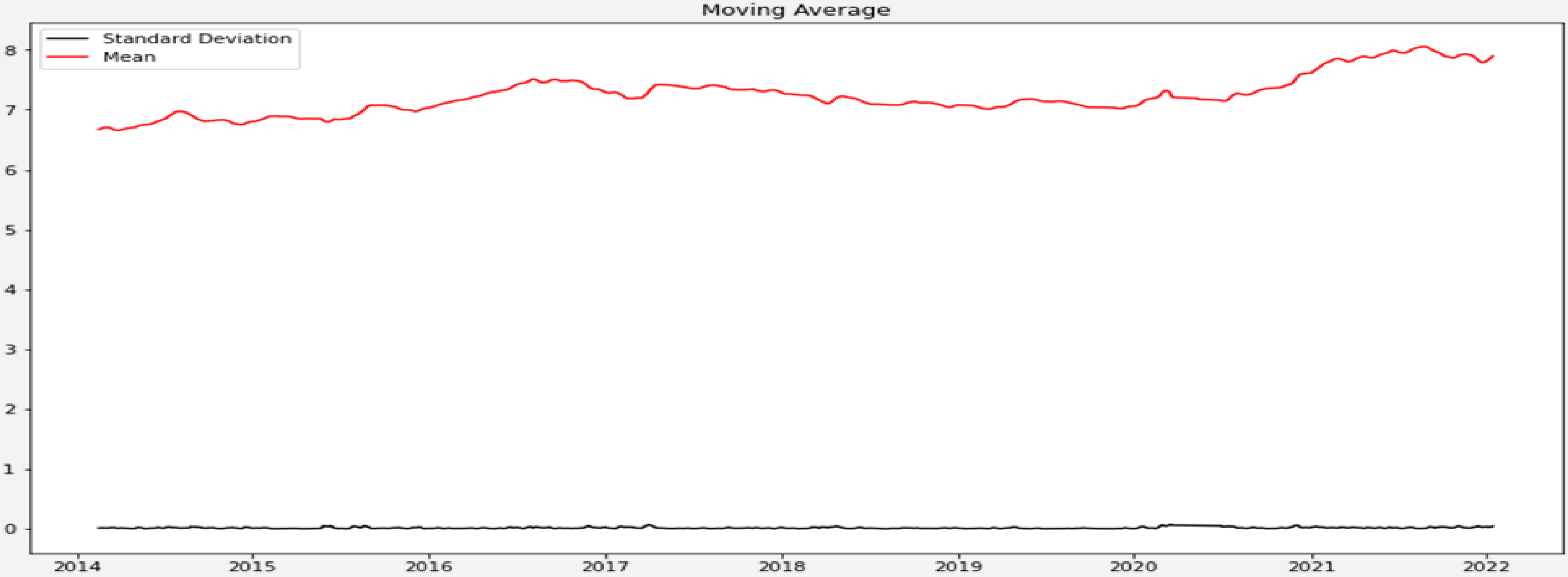
- In our case the series of data are not stationary so we decompose the seasonality and trend from our series before we can take a time series analysis.

- Hence, to decompose the series following steps are performed:
  I. The log of series is taken.



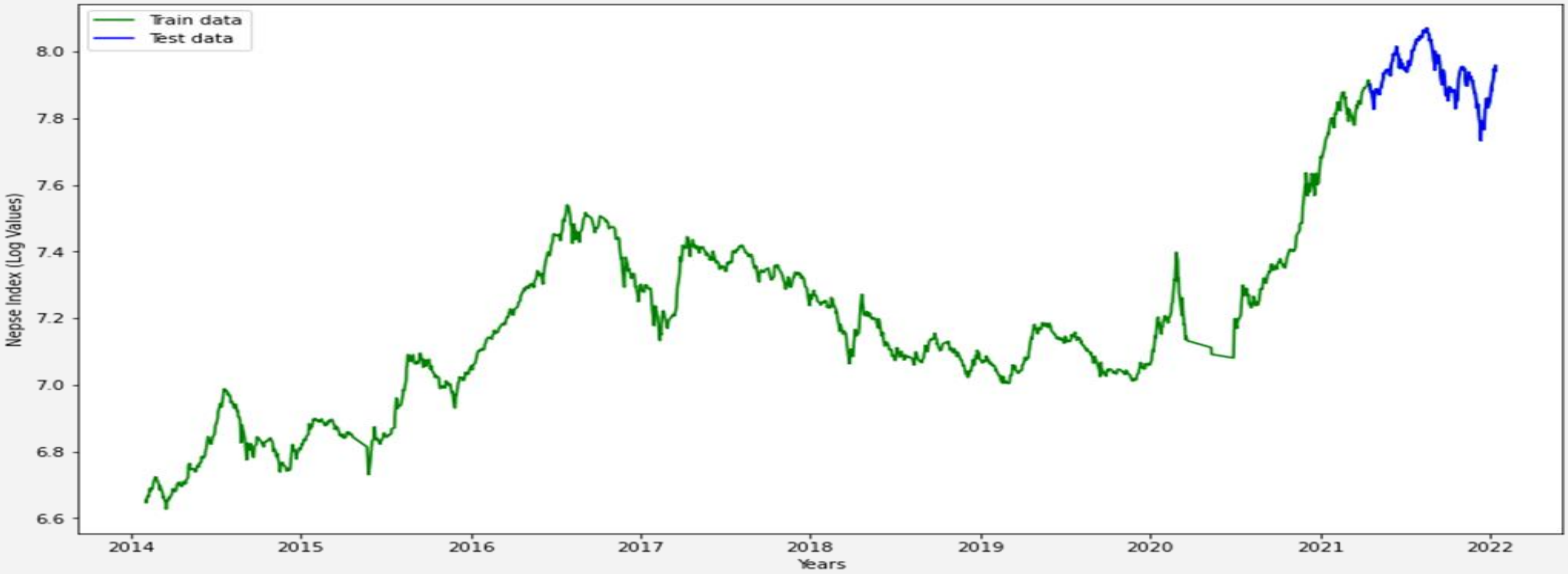
4.3.1 Stationarity Test – Eliminate Trend (Contd.)

II. Then we calculated the rolling average of the series. (Average of 12 days are taken).
III. Lastly, we calculate the mean consumption value at each subsequent point in the series.



4.4 Develop and Split Data

Now we'll develop an ARIMA model and train it using the stock's closing price from the series data. So, let's visualize the data by dividing it into training and test sets.



Note: Here we are using 90% of data for training and remaining 10% for testing.

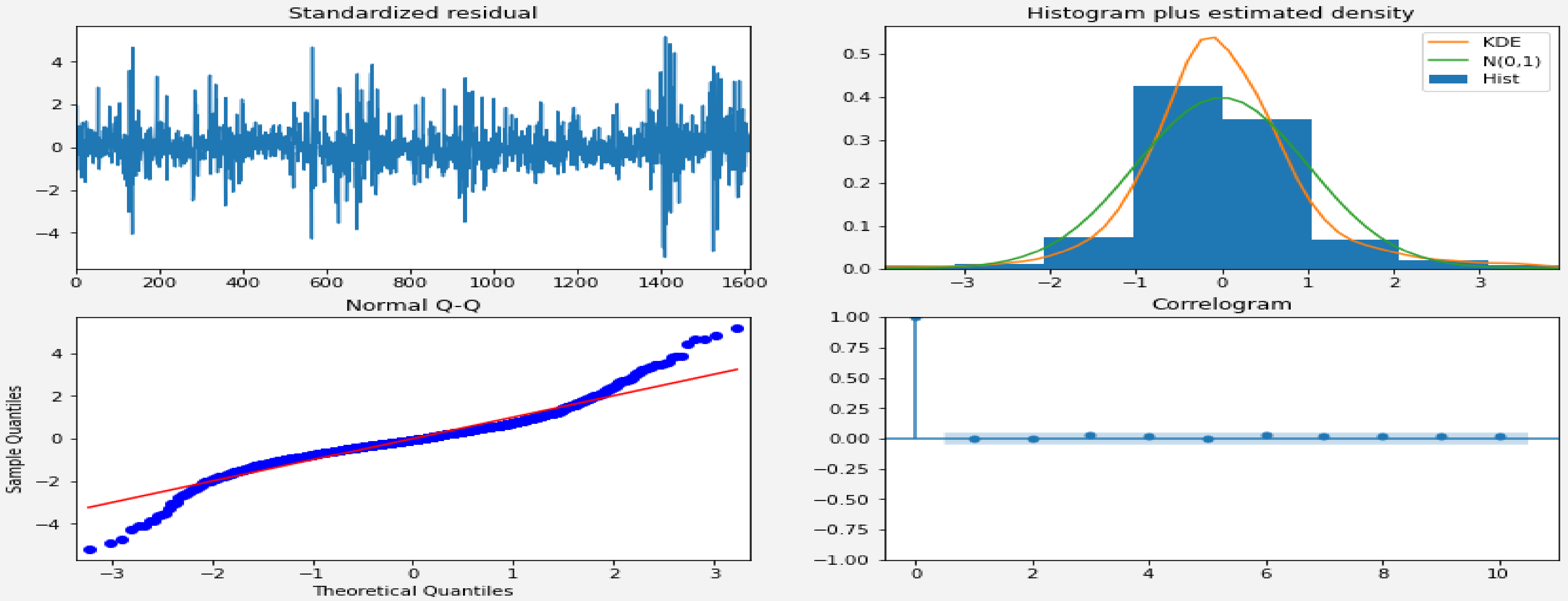
4.5 Discover Optimal Model (Auto ARIMA)

- auto_arima() uses a stepwise approach to search multiple combinations of p,d,q parameters and chooses the best model that has the least AIC (Akaike Information Criteria).

```
Performing stepwise search to minimize aic
 ARIMA(0,1,0)(0,0,0)[0] intercept   : AIC=-9563.437, Time=0.51 sec
 ARIMA(1,1,0)(0,0,0)[0] intercept   : AIC=-9608.863, Time=0.57 sec
 ARIMA(0,1,1)(0,0,0)[0] intercept   : AIC=-9610.389, Time=2.14 sec
 ARIMA(0,1,0)(0,0,0)[0]             : AIC=-9559.213, Time=0.35 sec
 ARIMA(1,1,1)(0,0,0)[0] intercept   : AIC=-9608.370, Time=1.58 sec
 ARIMA(0,1,2)(0,0,0)[0] intercept   : AIC=-9608.415, Time=1.09 sec
 ARIMA(1,1,2)(0,0,0)[0] intercept   : AIC=-9607.305, Time=3.55 sec
 ARIMA(0,1,1)(0,0,0)[0]             : AIC=-9607.745, Time=0.39 sec

Best model:  ARIMA(0,1,1)(0,0,0)[0] intercept
Total fit time: 10.215 seconds
                               SARIMAX Results
==============================================================================
Dep. Variable:                      y   No. Observations:                 1615
Model:               SARIMAX(0, 1, 1)   Log Likelihood                4808.194
Date:                Tue, 01 Feb 2022   AIC                          -9610.389
Time:                        17:38:20   BIC                          -9594.229
Sample:                             0   HQIC                         -9604.391
                               - 1615
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
intercept      0.0008      0.000      2.130      0.033    6.18e-05      0.001
ma.L1          0.1757      0.013     13.221      0.000       0.150      0.202
sigma2         0.0002   3.03e-06     49.863      0.000       0.000      0.000
===================================================================================
Ljung-Box (L1) (Q):                   0.00   Jarque-Bera (JB):              1291.53
Prob(Q):                              1.00   Prob(JB):                         0.00
Heteroskedasticity (H):               1.64   Skew:                             0.29
Prob(H) (two-sided):                  0.00   Kurtosis:                         7.34
===================================================================================
```

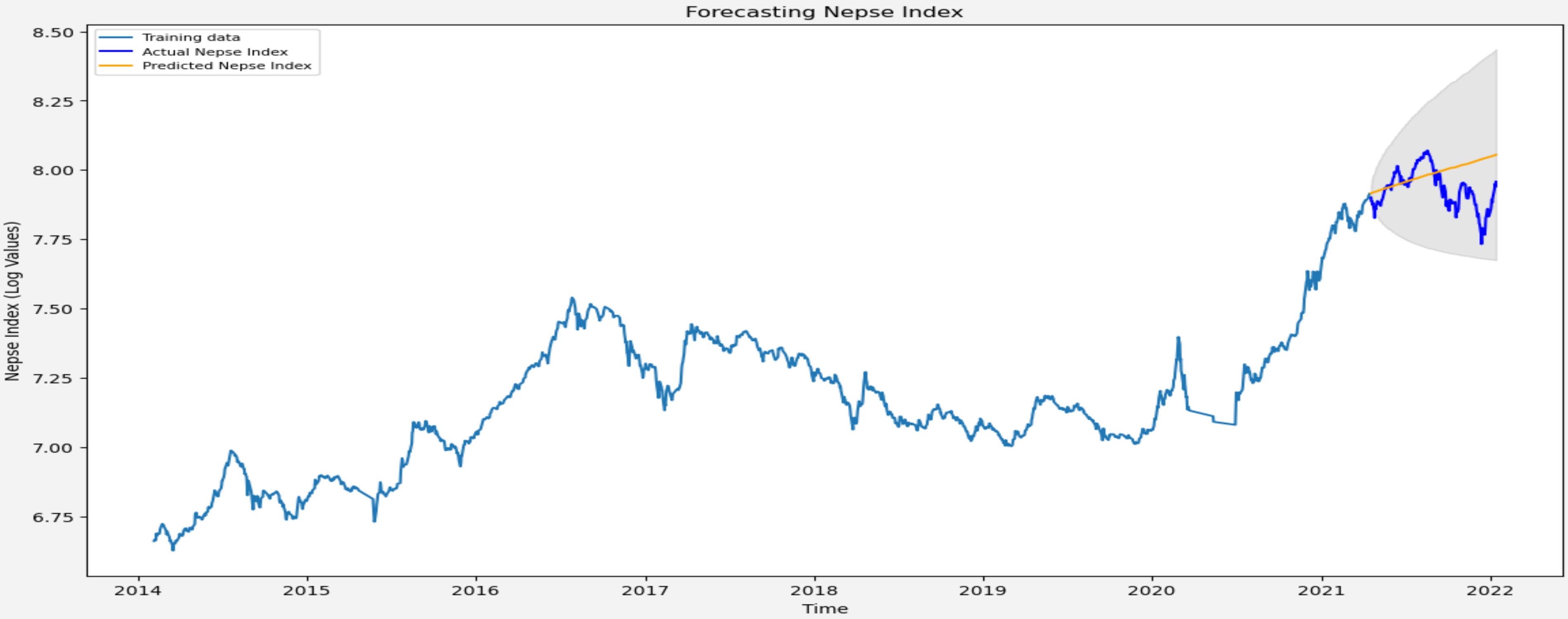
4.5.1 Interpret the residual plots in ARIMA model (Model Diagnostic)



➢ As a result, the Auto ARIMA model assigned the values 0, 1, and 1 to, p, d, and q, respectively.

4.6 Forecasting the Index

- Let's now begin forecasting stock prices on the test dataset with a 95% confidence level.



4.7 Evaluating Forecast Results and Report Performance

- Let’s take a look at some of the most common accuracy metrics for evaluating forecast results:

```
MSE: 0.011568610948811502
MAE: 0.08190843473354056
RMSE: 0.10755747741933845
MAPE: 0.010382474954984188
```

- From above the best forecast tool is MAPE because other error metrics are quantities. That implies, this tells us that the average deviation between the predicted value and the actual value. So, you can’t really use them to compare the forecasts of two different scaled time series.

- Hence, we obtained 0.0104 ~ = 1.04% i.e., MAPE implies the model is about 98.96% accurate in predicting the next 15 observations.

05

Findings and Conclusions

Findings and Conclusions

- The ARIMA model has been widely utilized in banking and economics since it is recognized to be reliable, efficient, and capable of predicting short-term stock market movements.

- Financial analysts, investors can use this prediction model to take trading decision by observing market behavior and sentiments.

- The method we have developed is simple and can be applied on any time series analysis.

- From the results, it can be observed that Mean Absolute Percentage Error (MAPE) is very small for the fitted model. Consequently, it shows that the model can relatively predict the time series data accurately.

- Hence, it is concluded that closing index of NEPSE in the current study shows a gradual increase for the upcoming trading days.

06

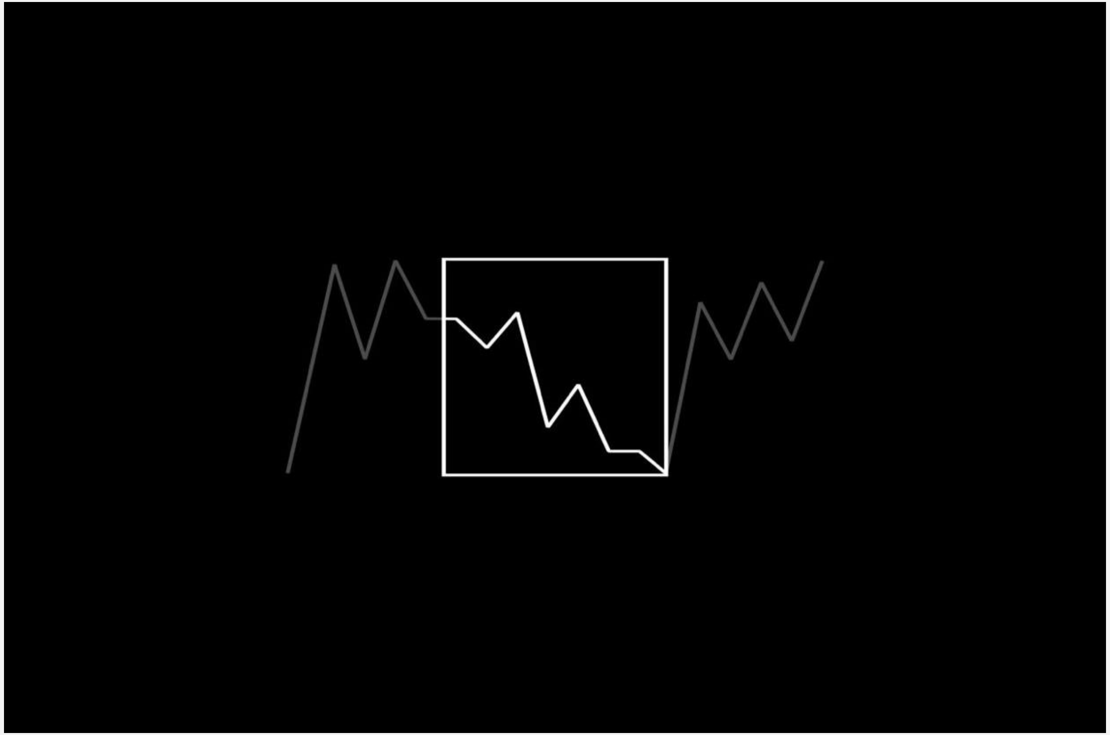
References

References

- (2019). In R. B. Paudel, K. J. Baral, R. R. Gautam, & S. B. Rana, Basic Finance (p. 397). Kathmandu: Asmita Books Publisher & Distributors (P) Ltd.
- Nepal Stock Exchange Ltd. (n.d.). Introduction. Retrieved from Nepal Stock Web site: http://www.nepalstock.com/about-us/introduction
- Levine, R., & Zervos, S. (1998). Stock Markets, Banks, and Economic Growth. The American Economic Review, 537-558.
- A, M. A., & K, S. K. (2017). Forecasting National Stock Price Using ARIMA Model. Global and Stochastic Analysis, 77-81.
- Gaire, H. N. (2017). Forecasting NEPSE Index: An ARIMA and GARCH Approach. NRB: Economic Review Article.
- Prasad, P. C., Manandhar, D., Maharjan, L., Rajkarnikar, L., & Jaiswal, A. (2018). Stock Market Forecast Using Time Series Analysis. 9th National Students' Conference on Information Technology (NaSCoIT), 9-13.
- C, M., M, V., & Chillale, N. R. (2019). Analysis of Daily Stock Trend Prediction Using ARIMA Model. International Journal of Mechanical Engineering and Technology (IJMET), 1772-1792.
- Adebiyi, A. A., Adewumi, A. O., & Ayo, C. K. (2014). Stock Price Prediction Using the ARIMA Model. UKSim-AMSS 16th International Conference on Computer Modelling and Simulation, 105-111.
- Hyndman, R. J., & Athanasopoulos, G. (2018). Forecasting: Principles and Practice. OTexts.
- Brownlee, J. (2017). Introduction to time series forecasting with python: how to prepare data and develop models to predict the future. Machine Learning Mastery.

“Failure Is The Frame, Not The Picture”

- Anonymous



Thank You!

Does anyone have any questions?