Final Project

Statistics 645-674: Time Series Forecasting

Course Instructor: Dr. Merrill Liechty

Group No: 7

Yong Hui Lin
Bharat Modhwadiya
Saloni Bhandari
Manoj Suchendra Reddy
Honey Jain (STAT-645-674)

INTRODUCTION

The dataset chosen for our time series forecasting project is the "BNS ((Bank of Nova Scotia) Bank Stock Market" dataset. This dataset contains historical data on the stock market performance of BNS Bank, a renowned financial institution. It includes information such as the opening price, high price, low price, closing price, and trading volume recorded over a specific period.

The selection of this dataset is based on several factors. Firstly, the dataset is highly relevant to our project as it directly relates to the stock market performance of BNS Bank. Analyzing and forecasting stock prices is of significant interest to investors, traders, and financial analysts, and this dataset provides an opportunity to address practical real-world scenarios.

Another reason for choosing this dataset is its availability and quality. The dataset has been obtained from Kaggle, a reputable platform for sharing and discovering datasets. This ensures that the data is reliable, accurate, and suitable for building forecasting models. Working with a high-quality dataset is crucial to ensure the validity and reliability of our forecasting results.

Furthermore, the dataset's time series nature makes it particularly suitable for time series forecasting. Time series data consists of observations recorded over time, where the order and time intervals between observations are essential. The BNS Bank dataset is structured as a time series, allowing us to leverage various time series forecasting techniques to capture patterns, trends, and seasonality in the stock market performance.

OBJECTIVE:

The objective of our time series forecasting project in using the "BNS Bank Stock Market" dataset is threefold. Firstly, we aim to develop an appropriate exponential smoothing model to forecast the future stock prices of BNS Bank accurately. Exponential smoothing is a widely used technique that accounts for trend, seasonality, and historical patterns in time series data. By applying this model, we seek to provide investors, traders, and financial analysts with reliable predictions to support their decision-making processes.

Secondly, we aim to build an ARIMA (AutoRegressive Integrated Moving Average) model to forecast the stock market performance of BNS Bank. ARIMA models are highly effective in capturing the complex dynamics of time series data, including autocorrelation, trend, and seasonality. By leveraging the ARIMA model, we intend to generate forecasts that account for these characteristics and provide valuable insights into the future direction of BNS Bank's stock prices.

Furthermore, we are hoping to also be able to compare the performances of the exponential smoothing model and ARIMA model to other simpler models such as time series linear regression model. We wanted to be able to see if these more advance and complex models can have a higher accuracy in comparison to more simple models such as linear regression.

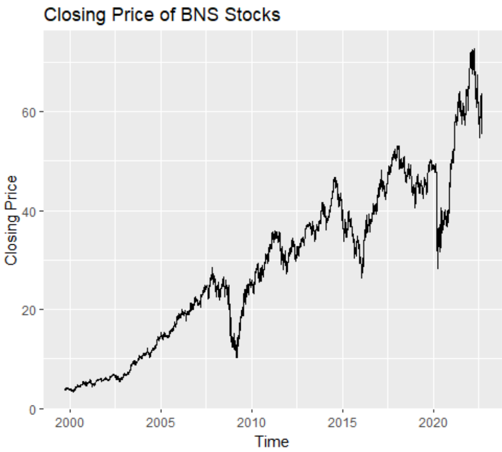
Ultimately, our objective is to deliver robust and accurate forecasting models that assist stakeholders in making informed investment decisions, managing risks, and optimizing their trading strategies. By analyzing the historical patterns and trends in the BNS Bank dataset, we aim to extract meaningful information that can guide financial decision-making and contribute to a better understanding of the factors influencing stock market behaviour. For simplicity, we will focus primary on the closing stock price of BNS as we conclude that closing stock price is a better indicator of the market value of the BNS stock market value, at least on that specific day.

EXPLOATORY DATA ANALYSIS (EDA):

Before proceeding with the modeling stage, we performed several data pre-processing steps to ensure the data's quality and consistency. These steps involved removing outliers, handling missing values through interpolation or imputation techniques, and checking for any underlying seasonality or trends. By cleaning and preparing the dataset appropriately, we aimed to ensure reliable and accurate forecasting results.

First, we would like to look at the overall closing stock price of BNS to see if there are abnormality within the dataset. From our initial observation, we clearly saw that there were dips in the graphs at around the years 2008, 2016, and 2020. We concluded that this might be because of a potential recession at that point in time, however, we do not have concrete evidence to support this conclusion. In addition, we did not saw any major gaps within the dataset, so the quality of the dataset is very good as there are no major gaps within the data.

Exhibit 1: EDA to gain deeper insights into the BNS Stock Price

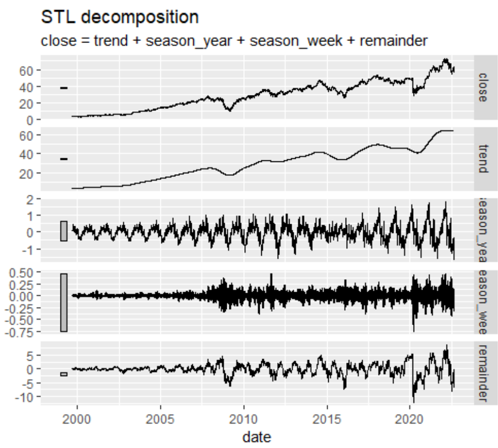


The line plot shows the closing prices of BNS stocks over time. The vertical axis represents the stock price in dollars, while the horizontal axis represents time from 2000 to 2022. The prices fluctuate over time, with a low of about $3 per share in 2000 and a high of about $70 per share in 2018. Overall, the plot shows a general positive trend of the stock prices of BNS over the past two decades, with no indication of any potential seasonality.

We wanted to further analysis the BNS stock price to ensure that there are no gaps within the data. Any potential gaps within the data would result in problems or errors later when performing analysis and modeling. After determining the difference between the timestamp of the indices, we noticed that there were implicit gaps with the data, which was expected as we knew that the stock market is only opens during business days (excluding weekends and holidays). As a result, we filled in the gaps of the date and utilized data interpolation to fill in the gaps of the data based on the linear model. We felt that since we will perform a box-cox transformation later due to trends in the dataset, the linear model would cause the less amount of variation within the data.

Once we filled in the gaps, the STL decomposition model has been used to analyze the seasonality and trends in the BNS bank stock price.

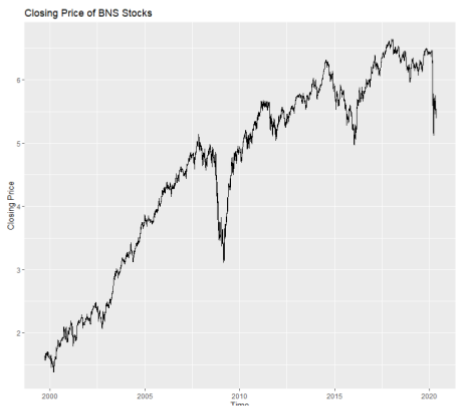
Exhibit 2: STL Decomposition of the BNS Closing Stock Price



The plot shows the results of the decomposition, where the "close" price of the BNS bank stock is decomposed into its trend, yearly seasonality, weekly seasonality, and remainder components. The y-axis shows the values of the different components, while the x-axis shows the timeline from 2000 to 2022. The plot also displays the BNS bank stock price data points over the years, with the values shown on the y-axis on the right side of the plot. The plot indicates that there is a clear trend but no clear seasonality pattern in the BNS bank stock price data over the years.

After we perform our EDA, we separated the dataset so that the first 90% of the dataset would be in the training data and the remaining 10% of the dataset would be the testing data. We also concluded that we need to do a box-cox transformation in order to normalize the data, so we determined that the ideal lambda value to use is 0.24.

Exhibit 3: The BNS Closing Stock Price after Box-Cox Transformation using Lambda = 0.24



The line plot shows the closing prices of BNS stocks over time. Similar to exhibit 1, we can clear see the dips in the same location however, the overall trend of the data is much more linear.

MODELING:

First, we wanted to create a baseline model that uses simpler models such as mean, STLF, naive, seasonal naive, DRIFT, and TSLM. We used these models to forecast the closing stock price of the testing data to determine how accurate each model is.

Exhibit 4: Accuracy table of the Forecasted BNS Closing Stock Price using simple models.

A tibble: 6 x 10

	.model	.type	ME	RMSE	MAE	MPE	MAPE	MASE	RMSSE
	<chr>	<chr>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1	Drift	Test	19.9	22.0	19.9	33.5	33.5	35.7	25.5
2	Mean	Test	29.4	31.4	29.4	50.4	50.4	52.6	36.5
3	Naive	Test	23.1	25.5	23.1	38.7	38.7	41.3	29.6
4	Seasonal_…	Test	21.8	24.3	21.8	36.3	36.4	39.0	28.2
5	stlf	Test	23.1	25.5	23.1	38.6	38.6	41.2	29.6
6	tslm	Test	-13.1	15.3	13.1	-27.8	27.8	23.3	17.8

From the accuracy, we can clearly see different accuracy indicators of each model. The model that has the lowest RMSE is the tslm model, which we would used in our final comparison as the baseline model.

Next, we wanted to create our exponential smoothing and ARIMA model. Prior to determining the best ARIMA model to use, we need to determine if the training data is stationary. As previously mentioned, there appears to be a trend within the data, so we know that the data is not stationary. We would need to determine both the number of first differences and seasonal differences that are necessary for the data to be stationary. Looking at the kpss_pvalue of the difference in the closing price of BNS, we determined that only a first difference is needed for the data to become stationary. Using the first differenced BNS stock price, we were ablet to look at both the ACF and PACF plots. From the plots, we see that there is an exponential decay in the PACF plot and that there is a significant spike at lag 3 in the ACF plot. An ideal ARIMA model that we could potentially use in the final comparison is ARIMA (0,1,3).

Once we obtained the ARIMA models that we would be using, we created models using our selected ARIMA model, the defaulted ARIMA model selected by RStudio (ARIMA (2,1,3)(0,0,1)[7] w/ drift), the defaulted ETS model, and the original TSLM model from our baseline.

Exhibit 5: Accuracy table of the Forecasted BNS Closing Stock Price

A tibble: 4 x 6

	.model	sigma2	log_lik	AIC	AICc	BIC
	<chr>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1	default	0.000642	17038.	-34061.	-34061.	-34005.
2	arima013	0.000644	17028.	-34045.	-34045.	-34010.
3	tslm	0.237	-5271.	-10866.	-10866.	-10804.
4	ets	0.000646	-5980.	11971.	11971.	12013.

Out of the four models chosen, we noticed that best model that we crated is the defaulted ARIMA model (ARIMA (2,1,3)(0,0,1)[7] w/ drift) which had the lowest AICc values. As a result, this would be our chosen model for forecasting the closing stock price of BNS.

Exhibit 6: Result of the Ljung-box analysis of the residual plot of BNS Closing Stock Price

A tibble: 1 x 3

	.model	lb_stat	lb_pvalue
	<chr>	<dbl>	<dbl>
1	ARIMA(box_cox(close, lambda))	72.4	0.000000144

Prior to forecasting, we also wanted to perform a Ljung-box test analysis to see if the residuals resemble white noise. As seen in exhibit 6, the lb-pvalue came back to be less than the critical value of 0.05, so we reject the null hypothesis concluding that the residuals does look like white noise.

FORECAST:

From the modeling step, we determined that the best model is the ARIMA (2,1,3)(0,0,1)[7] w/ drift. Using this model, we forecasted the closing stock price of BNS. Based on the accuracy of the model, we saw that the RMSE between the forecasted closing price and the testing model is 21.9.

Exhibit 7: Accuracy of the ARIMA (2,1,3)(0,0,1)[7] w/ drift on the Testing Data

```
# A tibble: 1 × 10
  .model       .type     ME  RMSE   MAE   MPE  MAPE  MASE RMSSE
  <chr>        <chr>  <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
1 ARIMA(box… Test    19.9  21.9  19.9  33.4  33.4  35.6  25.5
```

Exhibit 8: Forecasted Closing Stock Price of BNS

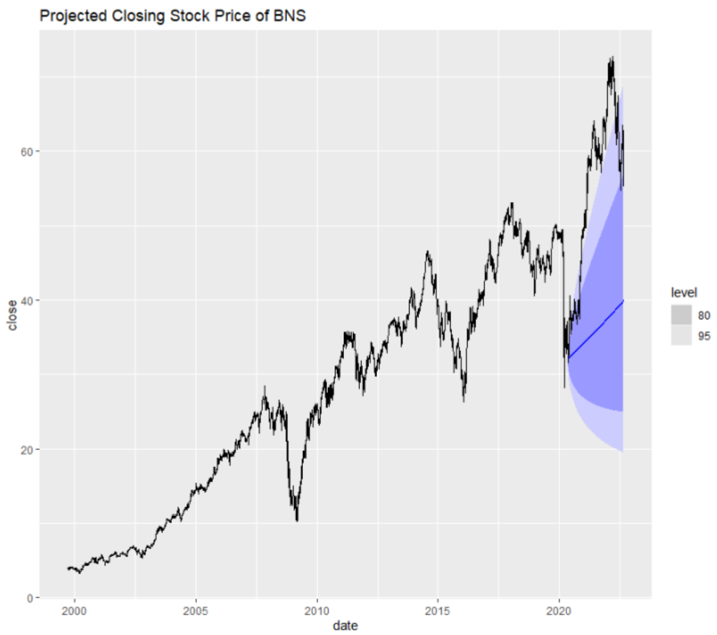
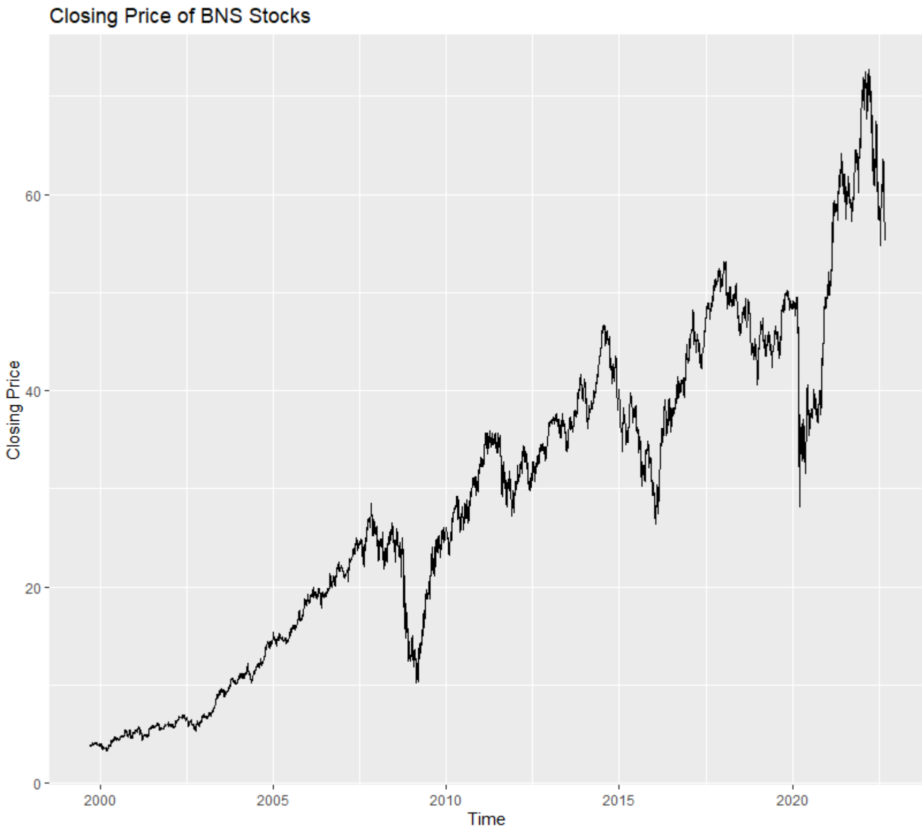


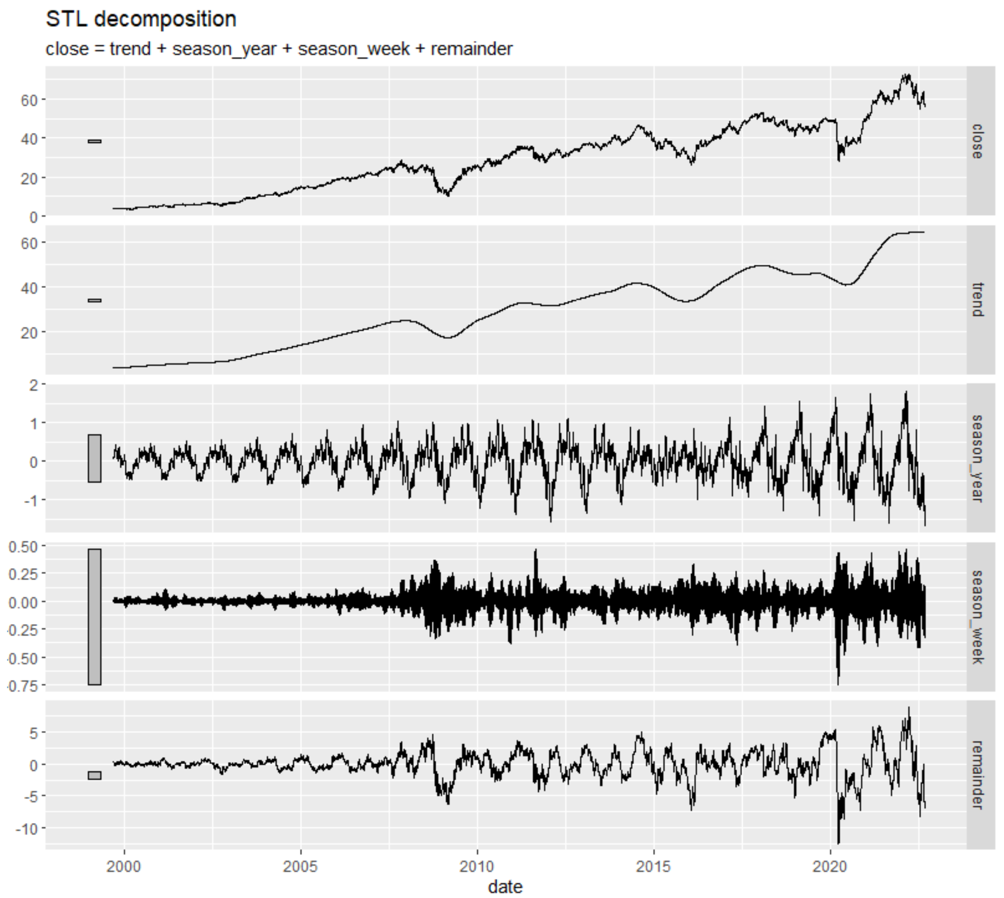
Exhibit 8 shows the forecasted closing stock price of BNS. The original black line presents the actual stock price history of BNS. The dark blue line represents the predicted point estimate of the closing stock price based on the ARIMA model. The regions around the line represents the different levels of confidence intervals. For the most part, wee can clearly see the actual closing stock price is generally outside of 95% confidence intervals so our predictions may not be as accurate as what we would like it to be.

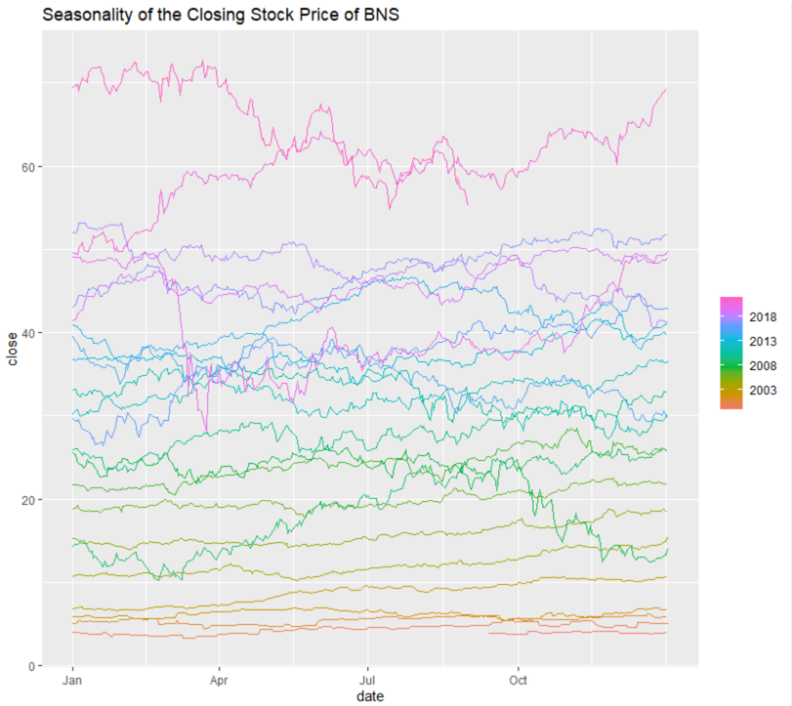
CONCLUSION:

The objective is to determine best model to use for forecasting the closing stock price of BNS. From our final candidates of potential models, we have chosen the defaulted ARIMA model (ARIMA (2,1,3)(0,0,1)[7] w/ drift) as this model yielded the lowest AICc values in comparison to the other candidate models. However, after we forecasted the testing data, we noticed that the actual closing stock price of BNS is actually slightly higher than the 95% confidence interval that the ARIMA model; thus, concluding that the accuracy of the forecast might not be as accurate as what we would like. However, we did notice that if we shorten the forecast to 30 days instead of 800+ days, the actual closing stock price of BNS lies just within the 95% confidence interval. As a conclusion, it is oftentimes can be relatively difficult to forecast stock prices over a long period of time since there are many different factors that could contribute to the changes or fluctuation of the market price, as shown here with the BNS stock.

APPENDIX:


Closing Price of BNS Stocks
Closing Price
60
40
20
0
2000
2005
2010
2015
2020
Time

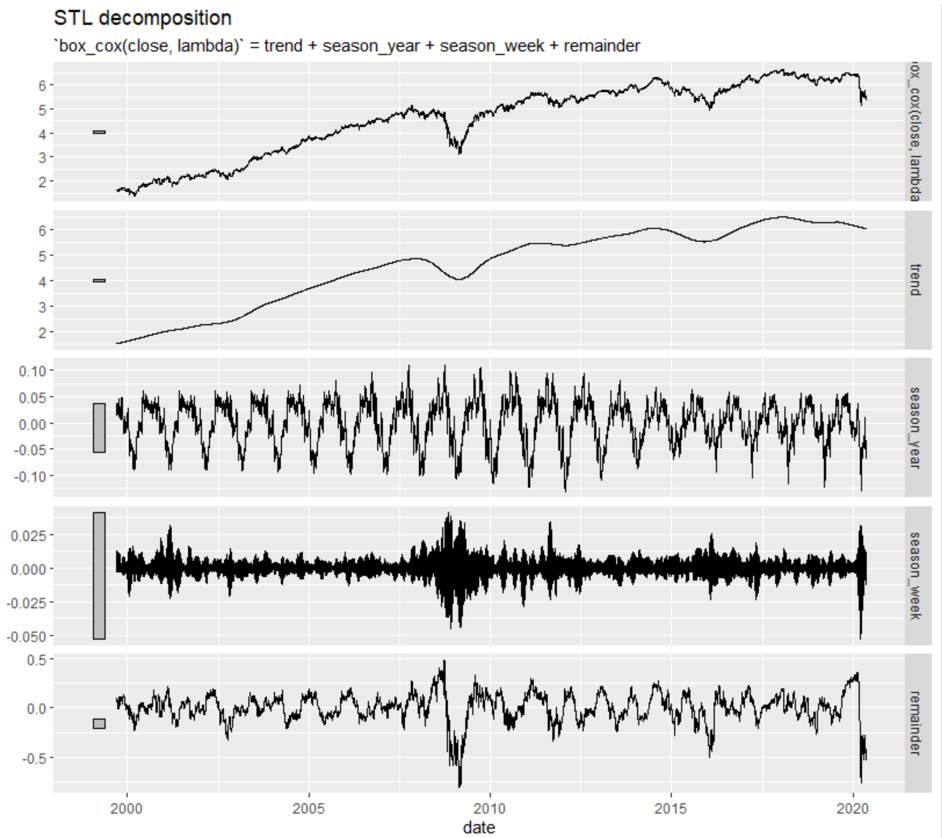

STL decomposition
close = trend + season_year + season_week + remainder
close
trend
season_year
season_week
remainder
2000
2005
2010
2015
2020
date

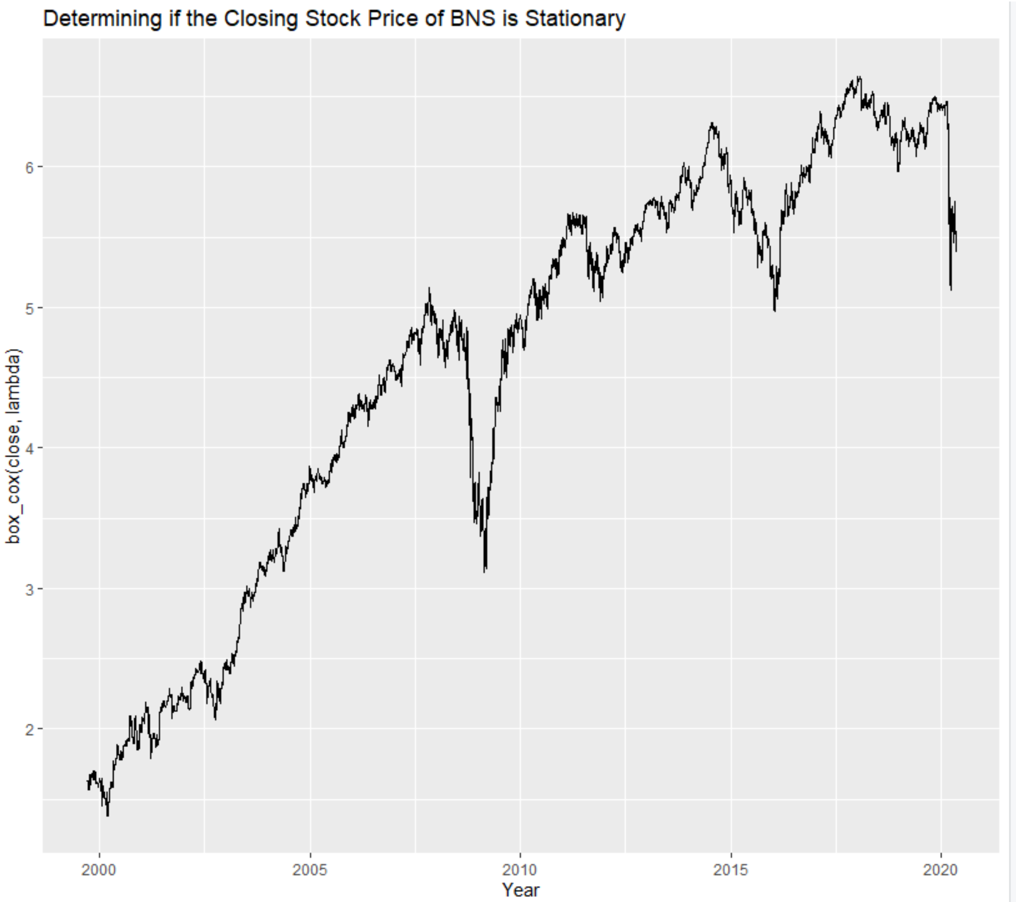


Strengths of Seasonal and Trend (F_t and F_s)

```
> stl_bns
# A tibble: 1 × 9
  trend_strength seasonal_strength_week
           <dbl>                  <dbl>
1           1.00                  0.180
```

STL Decomposition after Box_cox transformation using lambda = 0.24





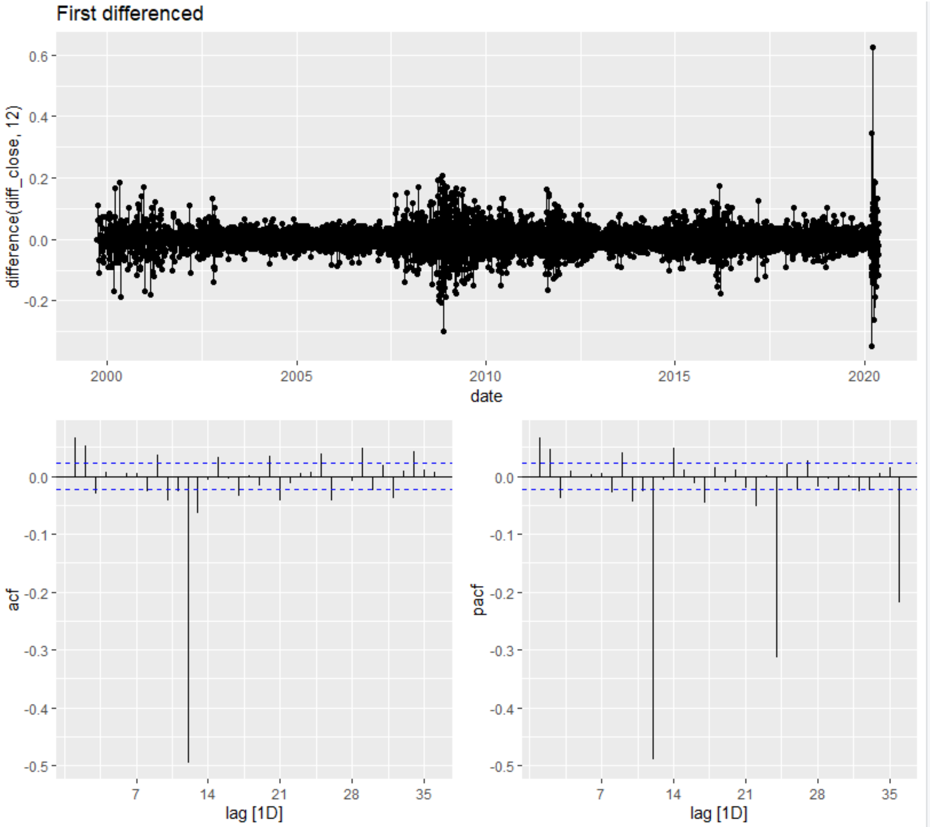
First Differences

```
> train_bns %>%
+   features(box_cox(close, lambda), unitroot_kpss)
# A tibble: 1 x 2
  kpss_stat kpss_pvalue
      <dbl>       <dbl>
1      56.9        0.01
> train_bns %>%
+   mutate(diff_close = difference(box_cox(close, lambda))) %>%
+   features(diff_close, unitroot_kpss)
# A tibble: 1 x 2
  kpss_stat kpss_pvalue
      <dbl>       <dbl>
1     0.291         0.1
```

Seasonal Differences

```
> train_bns %>%
+   mutate(season_close = box_cox(close, lambda)) %>%
+   features(season_close, unitroot_nsdiffs)
# A tibble: 1 x 1
  nsdiffs
    <int>
1       0
```

ACF and PACF plots generated from the First Differenced BNS Closing Stock Price



Residual Plot