Stationarity and Differencing

Kyla Ayop | Jolia Keziah Balcita

November 2, 2023

Contents

I. INTRODUCTION

In this paper, we will explore the essential concepts of stationarity and differencing of our given data. Stationarity is a key foundation in time series analysis as it serves as the basis for various statistical methods and models, and is critical for accurate forecasting and analysis.

Differencing, a key technique, is employed to achieve stationarity. This step is crucial as non-stationary data can lead to inaccurate results and predictions.

Our goal is to formulate a comprehensive analysis of the data's characteristics before and after differencing, providing clarity on the necessary transformations for an efficient time series analysis.

II. THE DATA

The data we will be using is from the historical stock prices of Google. More specifically, we will be analysing the closing prices of Google stocks in the year 2017. This data was recorded by Yahoo Finance.

```
library(fpp3)
library(urca)
```

```
gafa_stock
```

```
## # A tsibble: 5,032 x 8 [!]
## # Key:       Symbol [4]
##    Symbol Date        Open  High   Low Close Adj_Close    Volume
##    <chr>  <date>     <dbl> <dbl> <dbl> <dbl>     <dbl>     <dbl>
##  1 AAPL   2014-01-02  79.4  79.6  78.9  79.0      67.0  58671200
##  2 AAPL   2014-01-03  79.0  79.1  77.2  77.3      65.5  98116900
##  3 AAPL   2014-01-06  76.8  78.1  76.2  77.7      65.9 103152700
##  4 AAPL   2014-01-07  77.8  78.0  76.8  77.1      65.4  79302300
##  5 AAPL   2014-01-08  77.0  77.9  77.0  77.6      65.8  64632400
##  6 AAPL   2014-01-09  78.1  78.1  76.5  76.6      65.0  69787200
##  7 AAPL   2014-01-10  77.1  77.3  75.9  76.1      64.5  76244000
##  8 AAPL   2014-01-13  75.7  77.5  75.7  76.5      64.9  94623200
##  9 AAPL   2014-01-14  76.9  78.1  76.8  78.1      66.1  83140400
## 10 AAPL   2014-01-15  79.1  80.0  78.8  79.6      67.5  97909700
## # ... with 5,022 more rows
```

```
google_2017 <- gafa_stock %>%
  filter(Symbol == "GOOG", year(Date) == 2017)
```

III. STATIONARITY

Historical Plot

```
google_2017 %>%
  autoplot(Close) +
  labs(y = "Closing stock price ($USD)", x = "Month") + theme_bw()
```

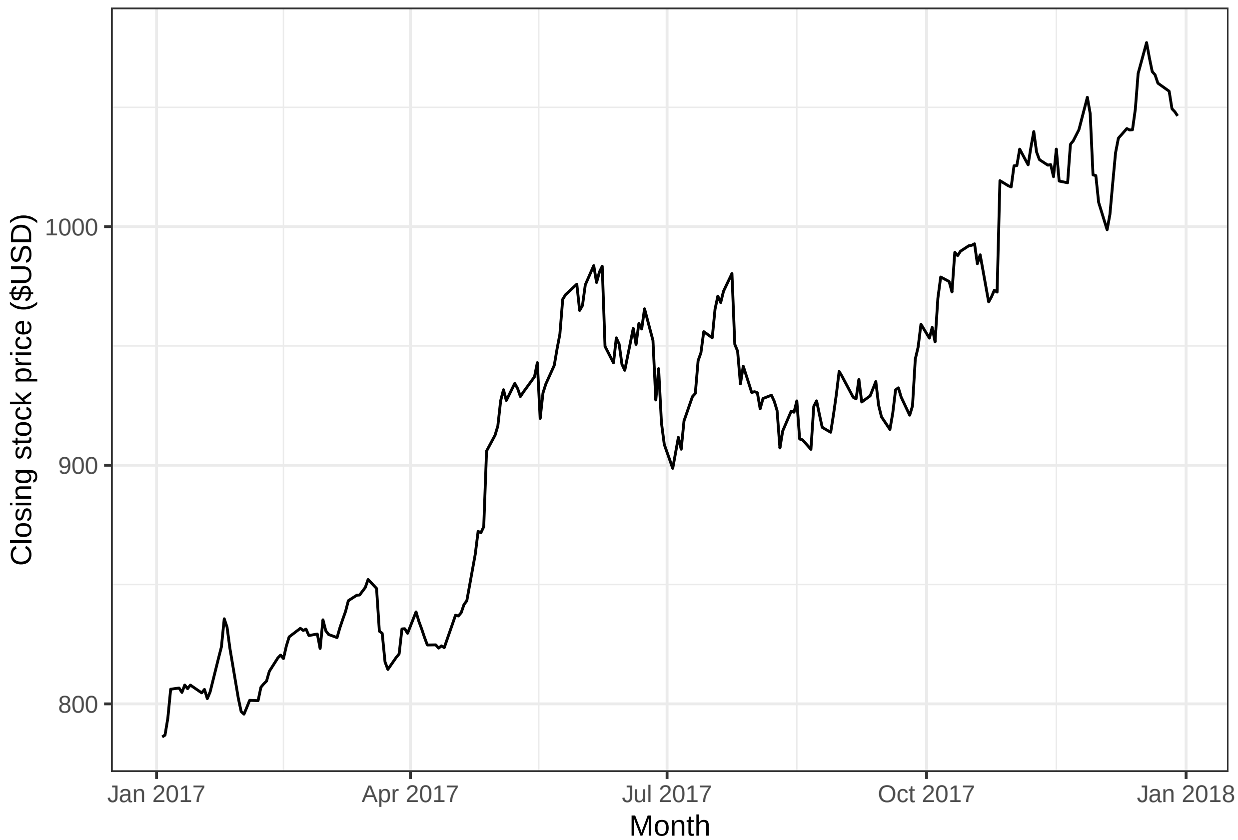


Figure 1: Closing Prices of Google Stocks in 2017.

The plot of the data reveals several important observations. Firstly, it depicts a clear upward trend over the course of the year, with a notable surge occurring in April to May. This suggests a consistent increase in the observed variable, particularly during that period.

Secondly, there appears to be no apparent seasonality which means that there are no recurring patterns or cycles that repeat at regular intervals throughout the year.

Taking these findings into consideration, it is concluded that the data is non-stationary. Additionally, to achieve accurate forecasting with this data, differencing should be performed. Differencing is a crucial step to transform non-stationary data into a stationary form, making it amenable to reliable and precise forecasting methods.

Autocorrelation

```
google_2017 %>%
  ACF(Close) %>%
  autoplot()+
  labs(y = "ACF", x = "Lags") + theme_bw()
```

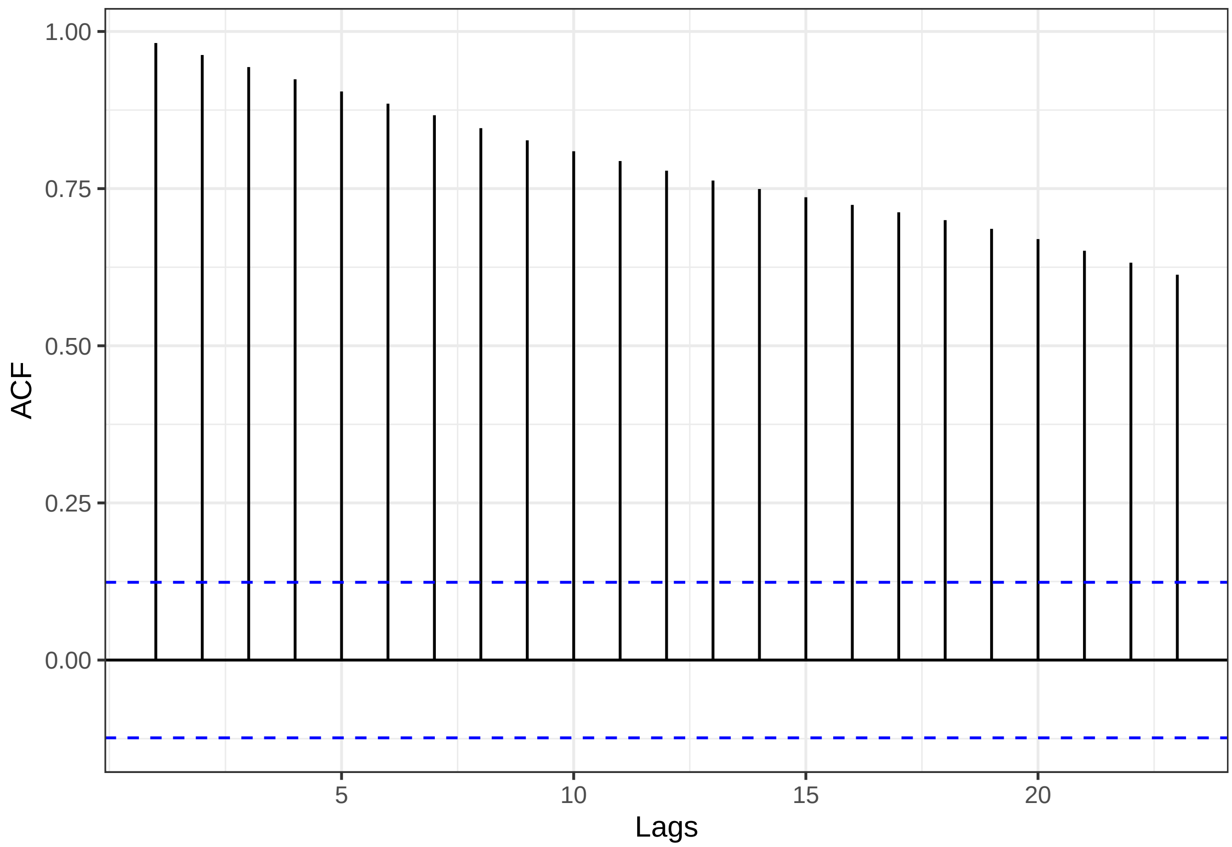


Figure 2: ACF Plot of Google Stocks' Closing Prices in 2017.

The ACF plot reveals a slow decrease in autocorrelation values which suggests that the data is non-stationary. With this, it is further proven that the data is non-stationary and differencing should be applied.

IV. DIFFERENCING

Since it has been observed that the original data is non-stationary from the time plot and ACF above, To address the non-stationarity of the data, the necessity of differencing with our data is highly needed. This is because the mean, variance, or other properties change over time. Thus, differencing is effective in making the series stationary by removing trends or seasonal patterns.

Applying Differencing

```
google_2017 %>%
  autoplot(difference(Close)) +
  labs(y = "Change in Google Closing Stock Price ($USD)", x = "Month") + theme_bw()
```

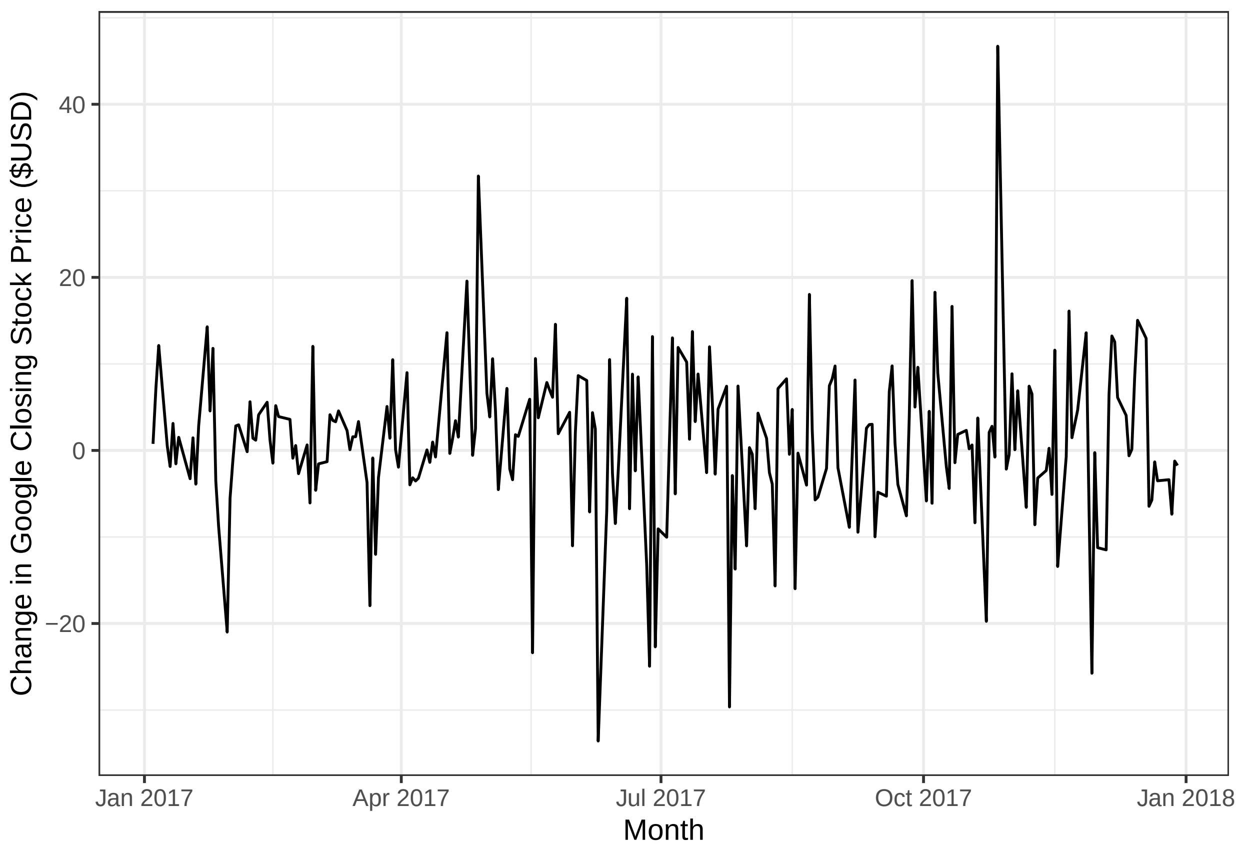
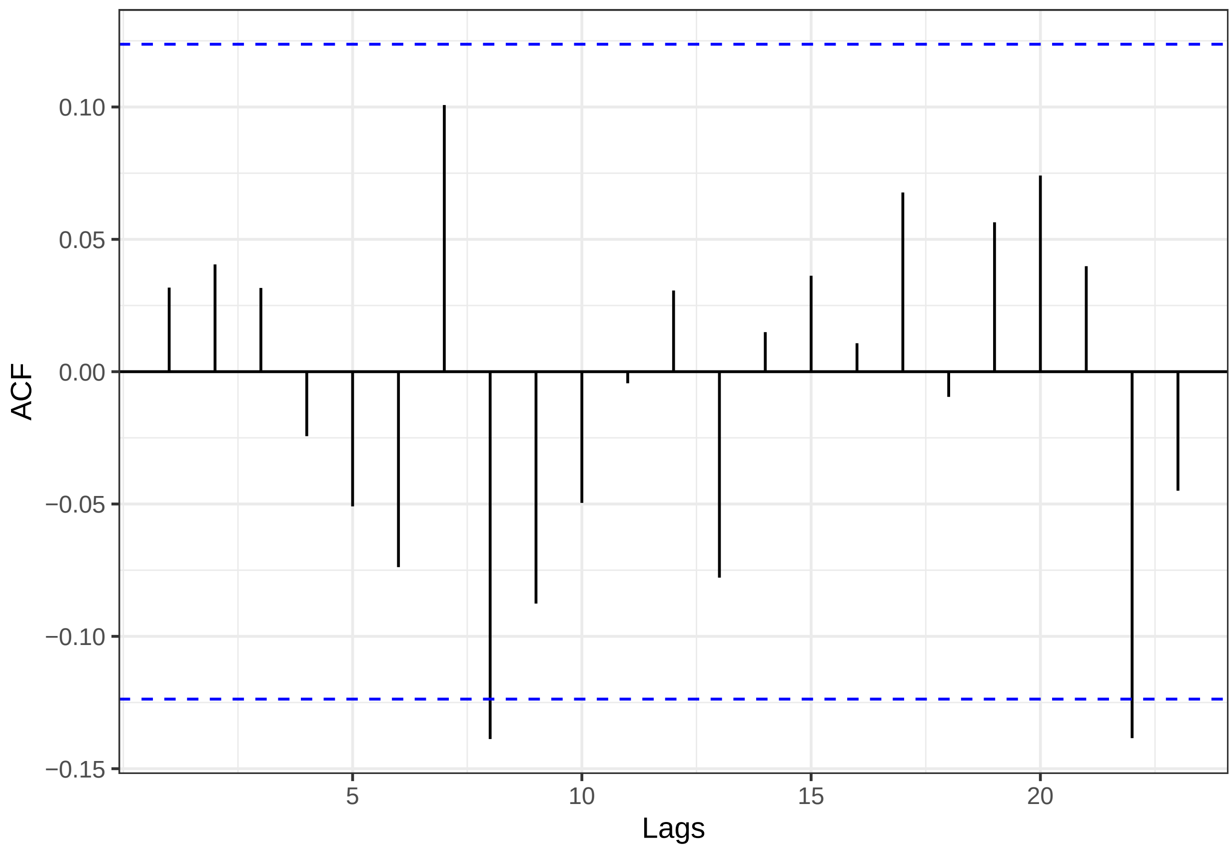


Figure 3: Closing Prices of Google Stocks in 2017 After Applying Differencing.

With the time plot of our differenced data, it has been observed that the series appeared to be stationary; there may be some high spikes at the end of the year. Nevertheless, the mean and variance are constant over time and is around change of zero.

ACF After Differencing

```
google_2017 %>%
  ACF(difference(Close)) %>%
  autoplot()+
  labs(y = "ACF", x = "Lags") + theme_bw()
```



As well as the time plot of the data, the ACF plot is also useful for identifying non-stationary time series. And with respect to our differenced Google stock close data for 2017, the ACF dropped to zero relatively quickly, thus being stationary, and it also looks just like that of a white noise series with only two autocorrelations outside of the 95% limits.

V. The Kwiatkowski–Phillips–Schmidt–Shin (KPSS) Test

Another way to determine more objectively whether differencing is required is to use a unit root test. In our analysis, we use the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test. In this test, the null hypothesis is that the series is stationary. Consequently, small p-values (e.g., less than 0.05) suggest that differencing is required, otherwise is stationary.

```
google_2017 %>%
  features(Close, unitroot_kpss)
```

```
## # A tibble: 1 x 3
##   Symbol kpss_stat kpss_pvalue
##   <chr>      <dbl>       <dbl>
## 1 GOOG        3.57        0.01
```

Here, with the original series of the Close price for the stock, a p-value of 0.01 has been obtained and with the significance level of 0.05, the p-value is lower than the significance level with the KPSS statistic is 3.565816. Therefore, we reject the null hypothesis of stationarity. This result tells us that the series is non-stationary.

This process of using a sequence of KPSS tests to determine the appropriate number of first differences is carried out using the unitroot_ndiffs() feature.

```
google_2017 %>%
  features(Close, unitroot_ndiffs)
```

```
## # A tibble: 1 x 2
##   Symbol ndiffs
##   <chr>   <int>
## 1 GOOG        1
```

As we saw from the KPSS tests above, one difference is required to make the Google 2017 series stationary.

```
google_2017 %>%
  mutate(diff_close = difference(Close) ) %>%
  features(diff_close, unitroot_kpss)
```

```
## # A tibble: 1 x 3
##   Symbol kpss_stat kpss_pvalue
##   <chr>      <dbl>       <dbl>
## 1 GOOG      0.0447         0.1
```

However, for the differenced series of the Close price for the stock the KPSS statistic is 0.04466562. The associated p-value is 0.1, which is greater than the significance level of 0.05. With this, the p-value is not less than the significance level. As a result, there is not enough evidence to reject the null hypothesis of stationarity, hence the differenced series is stationary.

VI. Augmented Dickey-Fuller Test

```
google_2017 <- gafa_stock %>%
  filter(Symbol == "GOOG", year(Date) == 2017) %>%
  select(Close)

# Perform ADF test using urca package
adf_result <- ur.df(google_2017$Close, type = "none", lags = 1)  # Adjust lag order as needed

# Print the ADF test results
summary(adf_result)
```

```
##
## ###############################################
## # Augmented Dickey-Fuller Test Unit Root Test #
## ###############################################
##
## Test regression none
##
##
## Call:
## lm(formula = z.diff ~ z.lag.1 - 1 + z.diff.lag)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -34.672  -4.282   0.188   4.799  45.733
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## z.lag.1    0.0010299  0.0006293   1.637    0.103
## z.diff.lag 0.0320641  0.0636804   0.504    0.615
##
## Residual standard error: 9.12 on 247 degrees of freedom
## Multiple R-squared:  0.01268,    Adjusted R-squared:  0.004689
## F-statistic: 1.587 on 2 and 247 DF,  p-value: 0.2067
##
##
## Value of test-statistic is: 1.6366
##
## Critical values for test statistics:
##       1pct  5pct 10pct
## tau1 -2.58 -1.95 -1.62
```

The google 2017 series for the closing price of the stock displays a clear increasing trend, hence we will use the 'trend' variation in the regression in the test. With the p-value of 0.104,it is higher than the 0.05 significance level. This result does not provide strong evidence to reject the null hypothesis (non-stationary) at the 5% significance level.Hence, based on the ADF test results, there's not enough evidence to reject the null hypothesis, indicating that the series is likely non-stationary.

```
google_2017 <- gafa_stock %>%
  filter(Symbol == "GOOG", year(Date) == 2017) %>%
  select(Close)
```

```r
# Take the differences of the Close values
google_2017_diff <- diff(google_2017$Close)

# Perform ADF test using urca package on the differenced data
adf_result_diff <- ur.df(google_2017_diff, type = "none", lags = 1)  # Adjust lag order as needed

# Print the ADF test results for the differenced data
summary(adf_result_diff)
```

```
##
## ###############################################
## # Augmented Dickey-Fuller Test Unit Root Test #
## ###############################################
##
## Test regression none
##
##
## Call:
## lm(formula = z.diff ~ z.lag.1 - 1 + z.diff.lag)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -33.907  -3.212   0.832   5.457  46.600
##
## Coefficients:
##            Estimate Std. Error t value Pr(>|t|)
## z.lag.1    -0.90702    0.08793 -10.315   <2e-16 ***
## z.diff.lag -0.05103    0.06360  -0.802    0.423
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 9.165 on 246 degrees of freedom
## Multiple R-squared:  0.4798, Adjusted R-squared:  0.4756
## F-statistic: 113.5 on 2 and 246 DF,  p-value: < 2.2e-16
##
##
## Value of test-statistic is: -10.3154
##
## Critical values for test statistics:
##       1pct  5pct 10pct
## tau1 -2.58 -1.95 -1.62
```

Now, with the differenced series, we have 'none' as the specific regression model for the new series since it will have no trend. With the p-value of <2.2e-16, it is extremely low (practically zero), well below the typical significance level of 0.05. With a very low p-value, the null hypothesis (non-stationary) is strongly rejected. This indicates that the series is stationary after differencing. Thus, indicating the successful removal of non-stationarity and allowing for more reliable time series analysis.

VII. References

Hyndman, R.J., & Athanasopoulos, G. (2021). *Forecasting: principles and practice, 3rd edition*, OTexts: Melbourne, Australia. OTexts.com/fpp3.

Lewis, R. (2021). Time Series Plots. *Study.com*. https://study.com/