Stationarity Analysis and Differencing of Apple Stock Data in 2014

Charlene P. Garridos & Ken Andrea Bahian

2023-11-05

Contents

1 R Libraries

```
library(urca)    # Unit Root and Cointegration Tests for Time Series Data
library(fpp3)    #"Forecasting: Principles and Practice" package
```

```
## -- Attaching packages ---------------------------------------------- fpp3 0.5 --

## v tibble      3.2.1     v tsibble     1.1.3
## v dplyr       1.1.2     v tsibbledata 0.4.1
## v tidyr       1.3.0     v feasts      0.3.1
## v lubridate   1.9.2     v fable       0.3.3
## v ggplot2     3.4.2     v fabletools  0.3.3

## -- Conflicts ------------------------------------------------- fpp3_conflicts --
## x lubridate::date()    masks base::date()
## x dplyr::filter()      masks stats::filter()
## x tsibble::intersect() masks base::intersect()
## x tsibble::interval()  masks lubridate::interval()
## x dplyr::lag()         masks stats::lag()
## x tsibble::setdiff()   masks base::setdiff()
## x tsibble::union()     masks base::union()
```

2 Time Series Data

```
apple_2014 <- gafa_stock %>%
  filter(Symbol == "AAPL", year(Date) == 2014)
head(apple_2014)
```

```
## # A tsibble: 6 x 8 [!]
## # Key:       Symbol [1]
##   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
```

3 Time Plot

```
# Time plot for 'Close' prices
autoplot(apple_2014, data = apple_2014, mapping = aes(x = Date, y = Close)) +
  labs(title = "Apple Stock Close Prices in 2014", x = "Date",
       y = "Change in Apple closing stock price($USD)")
```

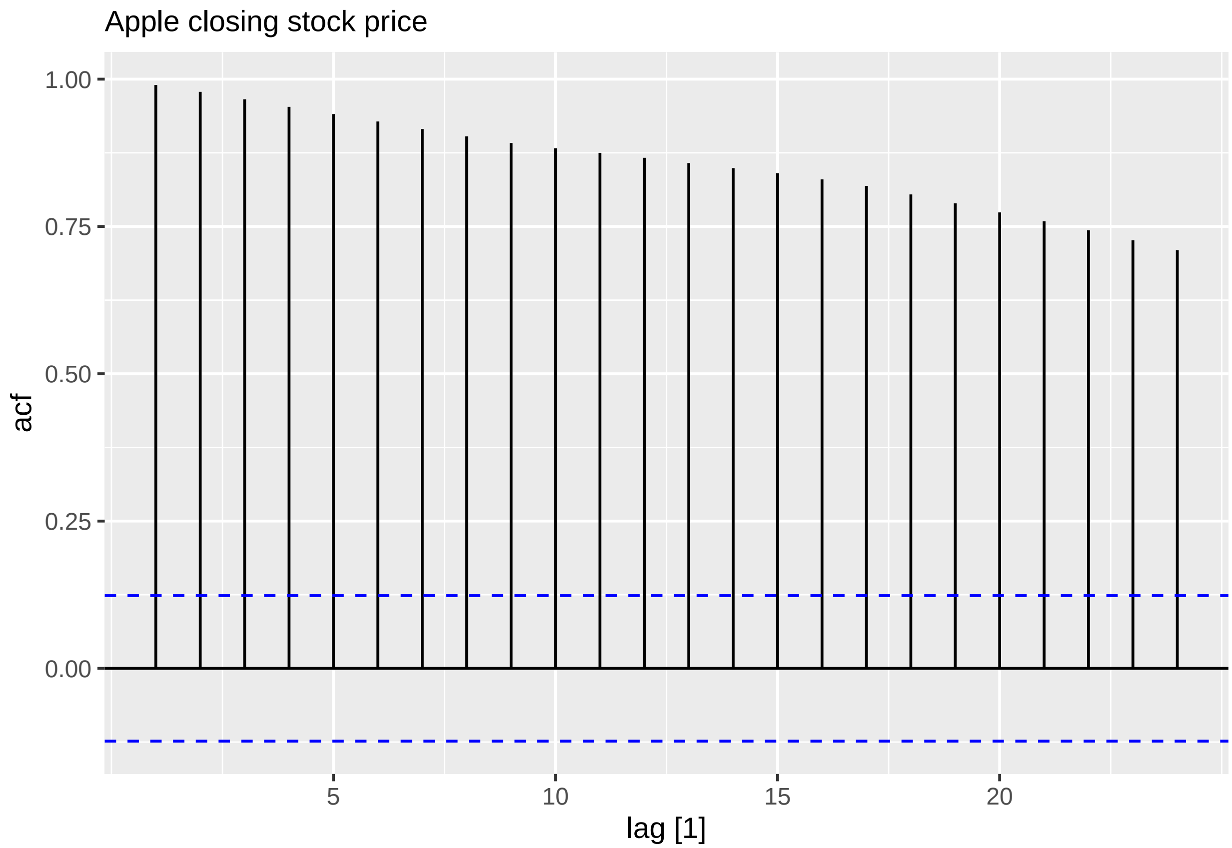
```
## Plot variable not specified, automatically selected `.vars = Open`
```

Interpretation

> Evidently, the plot may have covariance independent of time, and the mean increases over time, indicating that it is not constant. Additionally, it exhibits volatility, signifying that the variance is not also constant (heteroskedastic). Consequently, the series is non-stationary, necessitating a transformation to convert it from a non-stationary state to a stationary one. This transformation is commonly referred to as differencing.

4 Autocorrelation Function (ACF)

```
# ACF plot
apple_2014 %>% ACF(Close) %>%
  autoplot() + labs(subtitle = "Apple closing stock price")
```



Interpretation

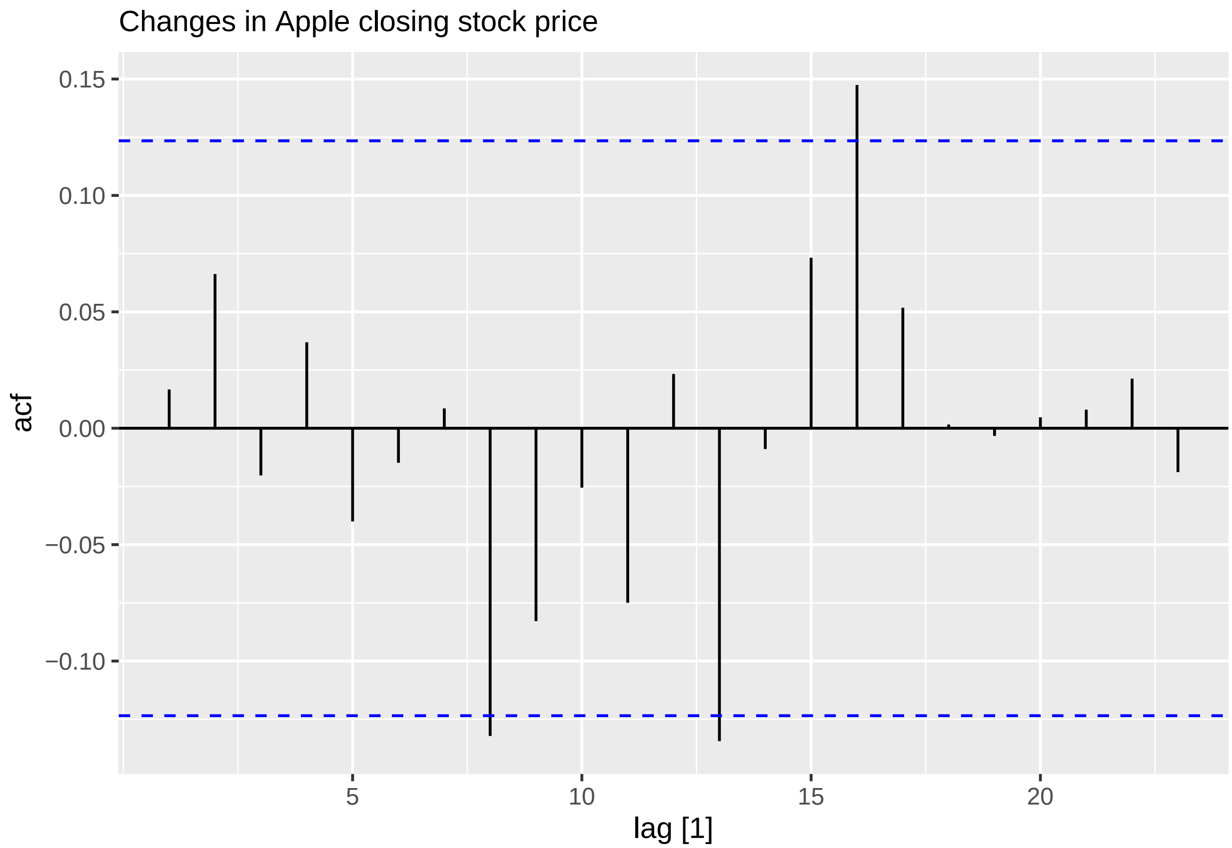
The lag is set to 1, and when analyzing the Autocorrelation Function (ACF) of the Apple closing stock price in 2014, it is evident that the ACF values do not drop to zero quickly. This lack of a rapid drop in ACF values implies non-stationarity in the data since the values decrease slowly over time.

Furthermore, it's important to note that the first ACF value is 1, indicating a strong positive correlation between the time series and its lagged values at a lag of 1. This suggests a pronounced one-period lag effect, signifying that each observation is strongly correlated with the preceding one.

Additionally, the vertical lines on the ACF plot exceed the horizontal blue lines which is the critical values. This phenomenon indicates significant autocorrelation in the data at the given lag, suggesting a potential pattern or dependence between the closing stock prices and their past values.

5 Differencing

```
# Differencing
apple_2014 %>% ACF(difference(Close)) %>%
  autoplot() + labs(subtitle = "Changes in Apple closing stock price")
```



Interpretation

Based on the observation, from a subjective perspective, it appears that the time series has become stationary after a single differencing or first differencing. Additionally, I've noted that three prominent spikes have exceeded the horizontal blue dashed lines, which serve as indicators of statistical significance. This suggests that the correlation at those specific lags is statistically significant, indicating the presence of a distinct pattern or dependence in the data at those time intervals.

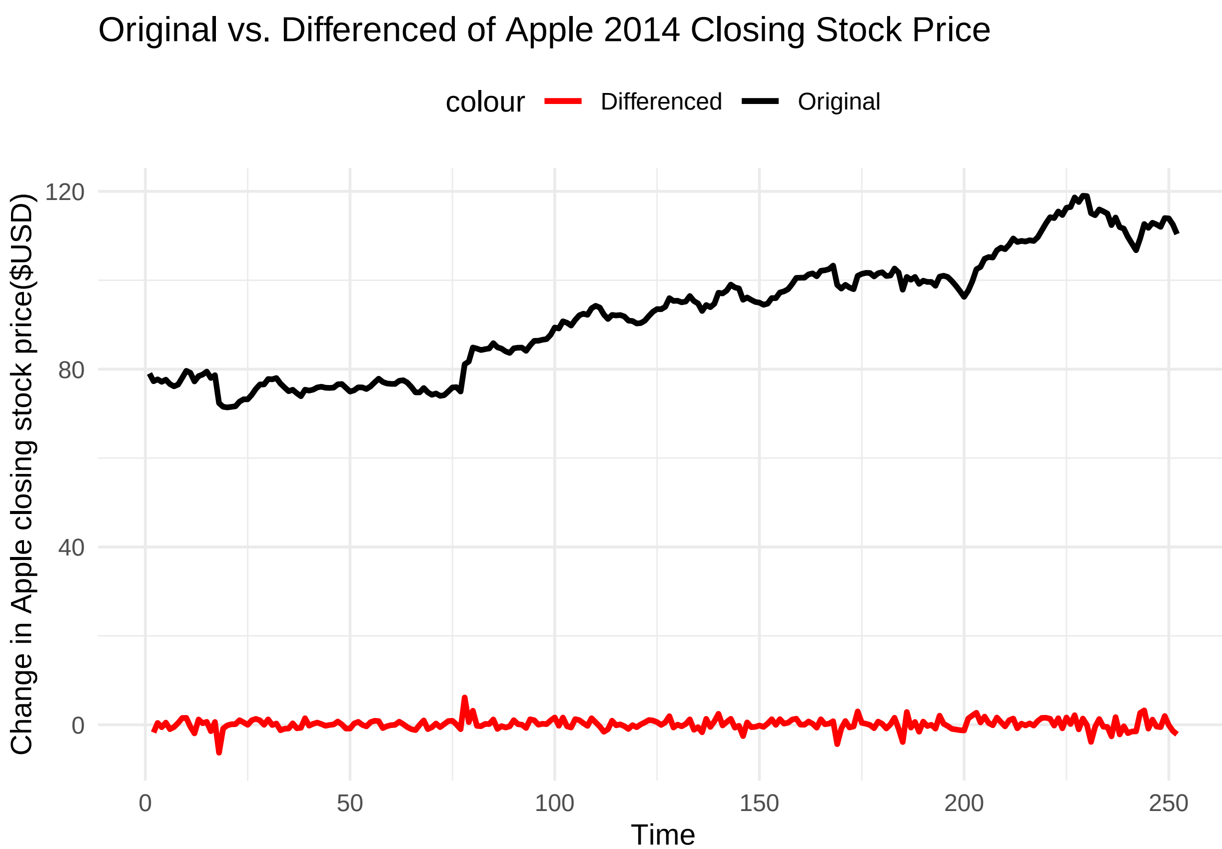
Conversely, a substantial portion of the autocorrelation values falls within the range defined by the blue lines. This implies that at those lags, the correlation is not statistically significant. In other words, the observed correlation might not differ significantly from what would be expected by random chance.

```r
#Compute the differenced series
diff_close <- diff(apple_2014$Close)

# Create a data frame for visualization
df <- data.frame(Time = seq_along(apple_2014$Close),
                 Original = apple_2014$Close,
                 Differenced = c(NA, diff_close))

# Create a line graph using ggplot2
ggplot(df, aes(x = Time)) +
  geom_line(aes(y = Original, color = "Original"), size = 1, linetype = "solid") +
  geom_line(aes(y = Differenced, color = "Differenced"), size = 1, linetype = "solid") +
```

```
  labs(title = "Original vs. Differenced of Apple 2014 Closing Stock Price",
       x = "Time", y = "Change in Apple closing stock price($USD)") +
  scale_linetype_manual(values = c("solid", "solid")) +
  scale_color_manual(values = c("Original" = "black", "Differenced" = "red")) +
  theme_minimal() +
  theme(legend.position = "top")
```



Interpretation

In the initial dataset analysis, we observed changing trends and fluctuations, making it unsuitable for standard statistical methods that assume data stability. To address this, we applied first-order differencing, effectively removing the overall trend with a single subtraction operation between consecutive values. Surprisingly, this simple step rendered the data stable, eliminating trends and systematic patterns. The resulting series showed steady mean, making it ideal for consistent analysis using statistical methods. This stability after just one difference simplified future analyses and provided a strong foundation for predictions.

6 Test H_0(Stationary) Using KPSS

```
apple_2014 %>%
  mutate(diff_close = difference(Close)) %>%
  features(diff_close, ljung_box, lag = 10)
```

```
## # A tibble: 1 x 3
##   Symbol lb_stat lb_pvalue
##   <chr>    <dbl>     <dbl>
## 1 AAPL      8.67     0.564
```

```
apple_2014 %>%
  features(Close, unitroot_kpss)
```

```
## # A tibble: 1 x 3
##   Symbol kpss_stat kpss_pvalue
##   <chr>      <dbl>       <dbl>
## 1 AAPL        4.07        0.01
```

Interpretation

> The LB statistic is 8.67, and its associated p-value is 0.564. Since the p-value (0.564) exceeds the common significance level of 0.05 (or 5%), you do not reject the null hypothesis (H0). This indicates a lack of significant autocorrelation in the residuals for lags up to 10, implying independence in the residuals up to that lag.
>
> On the other hand, the KPSS test is employed to ascertain whether a time series exhibits stationarity around a deterministic trend. The KPSS statistic is 4.07, and its associated p-value is 0.01. With a p-value of 0.01, which falls below the typical significance level of 0.05, we reject the null hypothesis (H0). This signifies that the time series is non-stationary, suggesting the presence of a unit root or a deterministic trend.

```
apple_2014 %>%
  mutate(diff_close = difference(Close)) %>%
  features(diff_close, unitroot_kpss)
```

```
## # A tibble: 1 x 3
##   Symbol kpss_stat kpss_pvalue
##   <chr>      <dbl>       <dbl>
## 1 AAPL      0.0781         0.1
```

Interpretation

> The KPSS statistic (0.0781) is lower than the critical values typically used at standard significance levels (e.g., 0.05 or 0.01), indicating that the time series does not exhibit a unit root or a deterministic trend. This suggests stationarity. In simpler terms, the series is stationary around its mean or displays trend-stationarity. The low KPSS statistic reinforces the effectiveness of differencing the time series (as we did with diff_close) in achieving stationarity.
>
> Moreover, the reported p-value is 0.1, which is higher than common significance levels, further supporting the conclusion of stationarity.
>
> In summary, the KPSS test results confirm stationarity in the differenced time series, and there is no compelling evidence of a unit root or non-stationary behavior. The null hypothesis is accepted, indicating stationarity after differencing.

```r
apple_2014 %>%
  features(Close, unitroot_ndiffs)
```

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

Interpretation

> After applying a single difference to the original closing prices of Apple stock for the year 2014, the resulting time series appears to become stationary. This suggests that differencing the data once has effectively removed significant trends or seasonality. Consequently, the series is now suitable for further time series modeling and analysis.
>
> In summary, a single difference (ndiffs = 1) in the 'Close' price series of Apple stock data from 2014, denoted as 'apple_2014,' is sufficient to achieve stationarity.

7 Bonus: Augmented Dickey-Fuller (ADF) test

```r
library(tseries)
```

```
## Registered S3 method overwritten by 'quantmod':
##   method            from
##   as.zoo.data.frame zoo
```

```r
# Perform Augmented Dickey-Fuller (ADF) test
adf_test <- ur.df(apple_2014$Close, type = "drift", lags = 1, selectlags = "AIC")

# Print ADF test results
summary(adf_test)
```

```
##
## ###############################################
## # Augmented Dickey-Fuller Test Unit Root Test #
## ###############################################
##
## Test regression drift
##
##
## Call:
## lm(formula = z.diff ~ z.lag.1 + 1 + z.diff.lag)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.4729 -0.6621 -0.0676  0.6961  5.9781
##
## Coefficients:
```

```
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.435436   0.551179   0.790    0.430
## z.lag.1     -0.003314   0.005920  -0.560    0.576
## z.diff.lag   0.019532   0.063902   0.306    0.760
##
## Residual standard error: 1.244 on 247 degrees of freedom
## Multiple R-squared:  0.001548,   Adjusted R-squared:  -0.006536
## F-statistic: 0.1915 on 2 and 247 DF,  p-value: 0.8258
##
##
## Value of test-statistic is: -0.5597 1.5096
##
## Critical values for test statistics:
##       1pct  5pct 10pct
## tau2 -3.44 -2.87 -2.57
## phi1  6.47  4.61  3.79
```

```
# Perform ADF test
result_adf <- adf.test(apple_2014$Close, alternative = "stationary")

# Print ADF test results
print(result_adf)
```

```
##
##  Augmented Dickey-Fuller Test
##
## data:  apple_2014$Close
## Dickey-Fuller = -3.2203, Lag order = 6, p-value = 0.08495
## alternative hypothesis: stationary
```

Interpretation

> We conducted two separate ADF tests using the urca and tseries libraries. In both cases, the obtained results exceeded the significance level (alpha = 0.05). Consequently, neither ADF test provided enough evidence to reject the null hypothesis. This suggests that the original data remains non-stationary.

```
#Compute the differenced series
diff_close <- diff(apple_2014$Close)

# Perform Augmented Dickey-Fuller (ADF) test on the differenced series
adf_test_diff <- ur.df(diff_close, type = "drift", lags = 1, selectlags = "AIC")

# Print ADF test results on the differenced
summary(adf_test_diff)
```

```
##
## ###############################################
## # Augmented Dickey-Fuller Test Unit Root Test #
## ###############################################
##
## Test regression drift
```

```
##
##
## Call:
## lm(formula = z.diff ~ z.lag.1 + 1 + z.diff.lag)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.3191 -0.6458 -0.0344  0.7455  6.0391
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.11926    0.07984   1.494    0.137
## z.lag.1     -0.91476    0.09014 -10.149   <2e-16 ***
## z.diff.lag  -0.06734    0.06393  -1.053    0.293
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.244 on 246 degrees of freedom
## Multiple R-squared:  0.4897, Adjusted R-squared:  0.4856
## F-statistic: 118.1 on 2 and 246 DF,  p-value: < 2.2e-16
##
##
## Value of test-statistic is: -10.1487 51.5058
##
## Critical values for test statistics:
##       1pct  5pct 10pct
## tau2 -3.44 -2.87 -2.57
## phi1  6.47  4.61  3.79
```

```r
# Perform ADF test on the differenced series
result_diff <- adf.test(diff_close, alternative = "stationary")
```

```
## Warning in adf.test(diff_close, alternative = "stationary"): p-value smaller
## than printed p-value
```

```r
# Print ADF test results for the differenced series
print(result_diff)
```

```
##
##  Augmented Dickey-Fuller Test
##
## data:  diff_close
## Dickey-Fuller = -5.7325, Lag order = 6, p-value = 0.01
## alternative hypothesis: stationary
```

Interpretation

> By employing two different types of Augmented Dickey-Fuller (ADF) tests initially, both tests failed to provide sufficient evidence to reject the null hypothesis. However, once we applied a first-order differencing to the original data before conducting the ADF test, the transformed data yielded significant evidence to reject the null hypothesis.

In the ADF test performed using the tseries library, we obtained a p-value of 0.01 for the differenced data. The p-value (0.01) is less than the typical significance level of 0.05. Therefore, you should reject the null hypothesis that the differenced series is non-stationary. The results suggest that the differenced series is likely stationary, which is a common assumption for time series analysis.

In summary, based on these ADF test results for the differenced series, we can reject the null hypothesis in favor of stationarity.