Modeling non-stationarity and finding an equilibrium

1. Definition

1. Augmented Dickey-Fuller (ADF) Test

The ADF test is a type of unit root test that checks for the presence of a unit root in a time series sample. The ADF equation for a time series Y_t is:

$$Y_t = \ + t + Y_{t-1} + {}^{p}_{i=1}{}_{i} Y_{t-i} + {}_{t}$$

Where:

- Y_t = The change in the series at time $t(Y_t - Y_{t-1})$.
- α = A constant term.
- βt = The coefficient for a time trend.
- γ = The coefficient for the lagged level of the series.
- δ_i = The coefficients for the lagged changes of the series.
- ε_t = Error term at time t.
- p = Number of lagged changes included.

2. Kwiatkowski-Phillips-Schmidt-Shin (KPSS) Test

The KPSS test is used to test for the stationarity of a time series around a deterministic trend.

$$Y_t = Y_{t-1} + {}_{t}, where_t = \ + t + {}^{t}_{i=1} u_i$$

Where:

- Y_t = The time series at time t.
- ρ = A coefficient indicating the autoregressive nature.
- μ = Constant term.
- τ = Coefficient for a deterministic trend.
- u_i = Random error terms that are stationary.

2. Description

When we're looking at time series data, sometimes the data doesn't play by the rules – it doesn't stick to a steady average or pattern. This kind of data is called "non-stationary." Think of "equilibrium" as the normal state of this kind of data. It's like finding the usual spot where the data likes to hang out.

To understand this better, we use special tests like the Unit Root test Augmented Dickey-Fuller (ADF) and the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test. These tests are like detective tools that help us determine the nature of the data (stationary or non-stationary) (Enders, 2014).

Once we know what kind of data we're dealing with, we can choose the right tools, like ARIMA, SARIMA, or ECM models, to make sense of it. Sometimes, we need to tweak the data a bit to make it more uniform and predictable – this is called making it "stationary." Doing this is important because it sets the stage for us to find out the data's "equilibrium" or its long-term trends.

3. Demonstration

The time series data is Apple stock's price for 10 years (2013-2023). First, I check the data to see if it is stationary or not by using Unit root test and KPSS test.

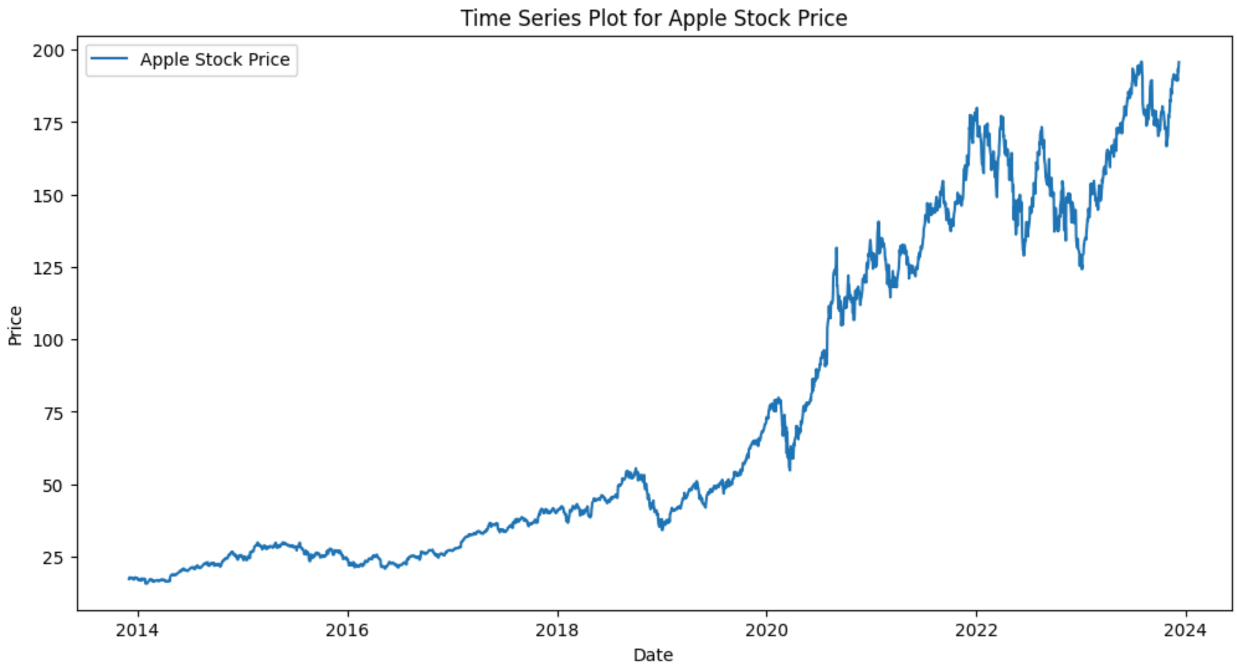
By running the Unit root test (ADF) the result is:

```
ADF Statistic: 0.4887409065312726
p-value: 0.984526837746281
Fail to reject the null hypothesis. The time series is likely
non-stationary.
```

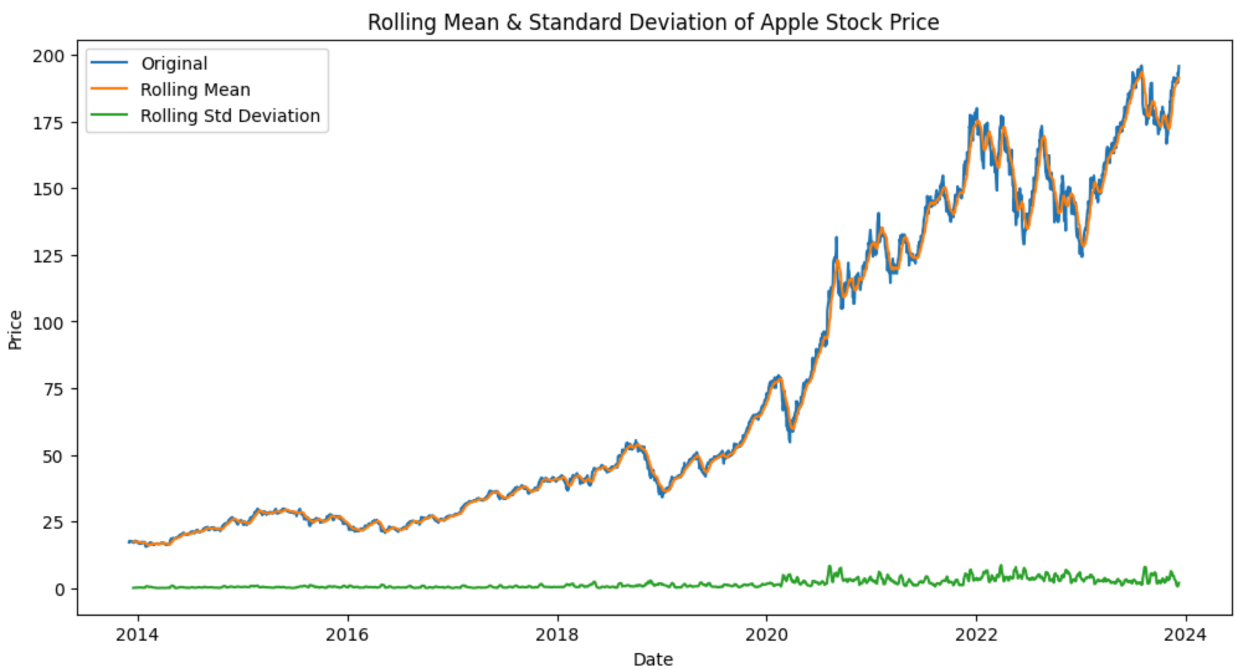
By running the KPSS test the result is:

```
KPSS Statistic: 7.405961395595039
p-value: 0.01
Reject the null hypothesis. The time series is likely non-stationary.
```

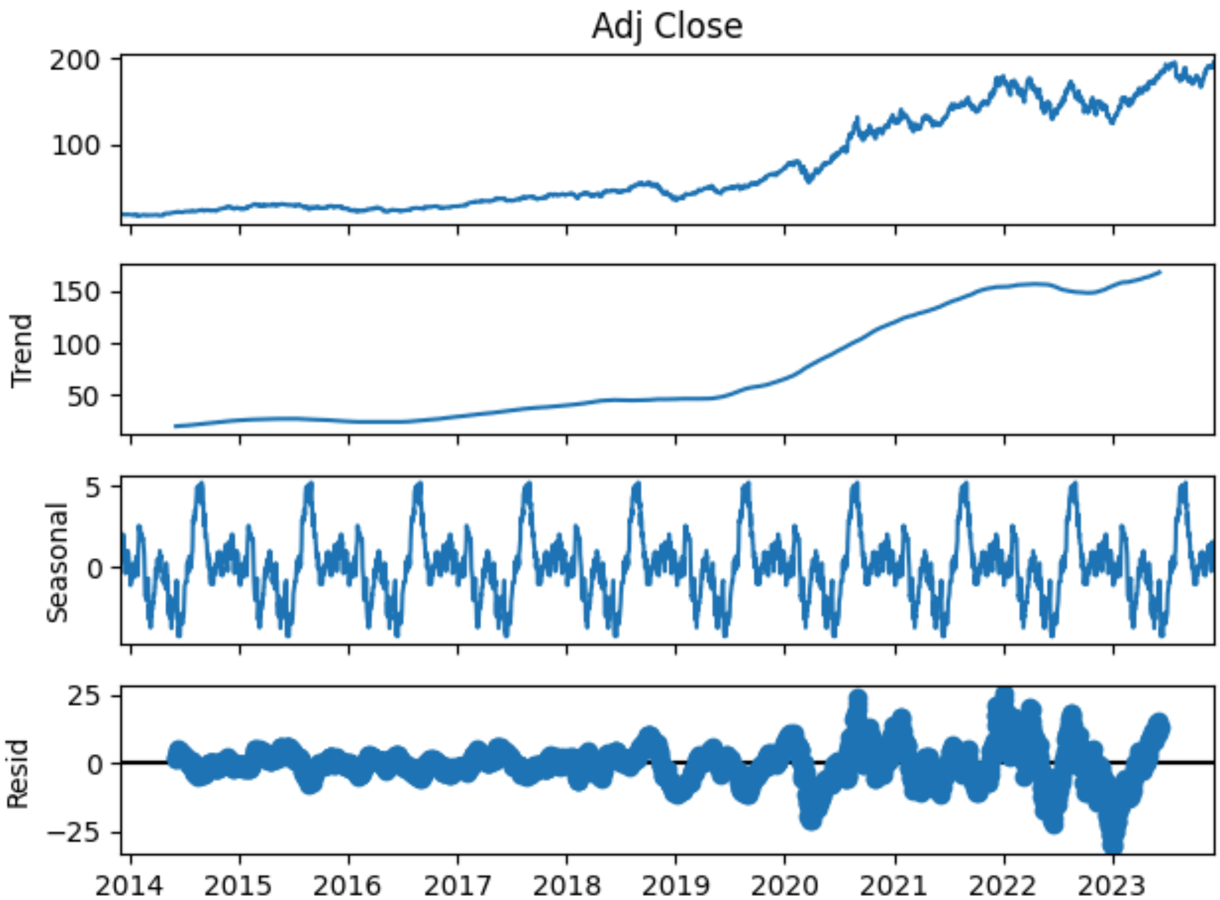
4. Diagram



Time Series Plot: There's a clear upward trend indicating that the stock price has been increasing over time. There are also some fluctuations within the trend, but the overall direction is upwards. This kind of trend is typical for growing companies in the tech sector, but it also suggests that the raw data is non-stationary because the mean price is not constant over time.



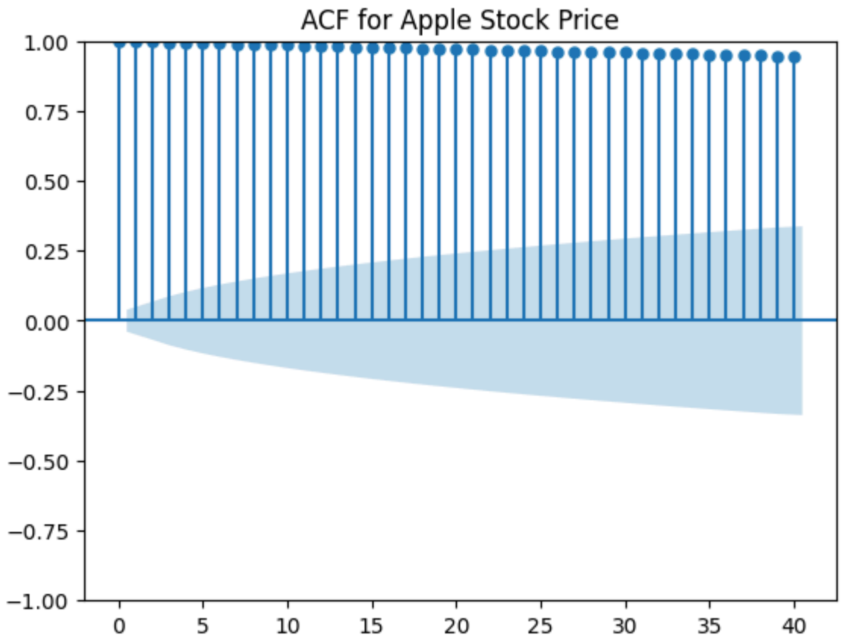
Rolling Statistics Plot: The rolling mean appears to be increasing, which confirms the non-stationarity we suspected from the time series plot. The rolling standard deviation is relatively stable but shows some variation, indicating changes in the volatility of the stock price over time. This plot reinforces the conclusion that the data is non-stationary since the mean and volatility are not constant.



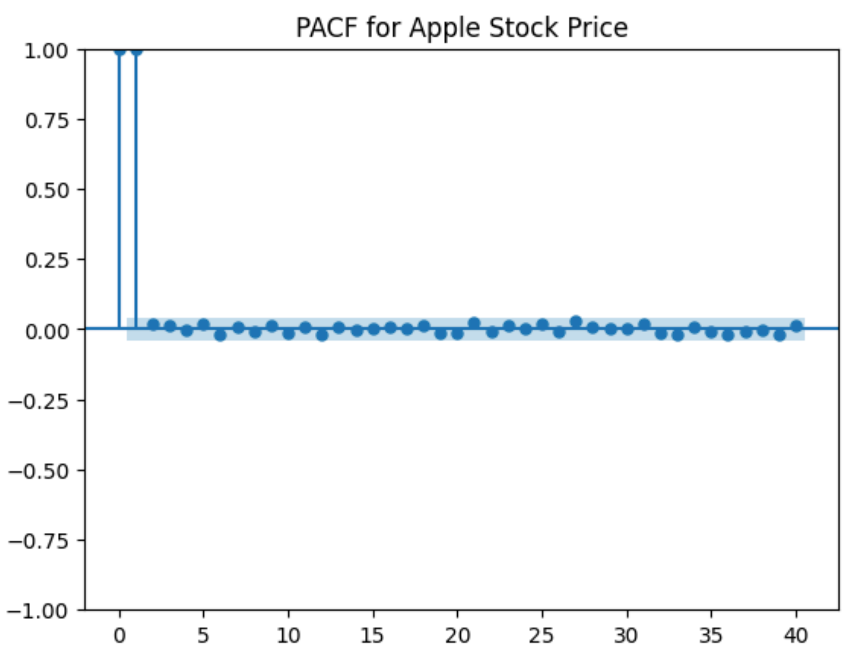
Decomposition Plot: The trend component clearly shows an upward trajectory, which is consistent with the time series and rolling statistics plots. The seasonal component doesn't show a clear seasonal pattern, which could suggest that the stock price doesn't have a strong seasonal cycle, or the cycle might not be captured due to the noise in the data. The residual (or irregular) component shows the random fluctuations after accounting for the trend and seasonality, which appear to be quite volatile and suggest that even after removing the trend, the data still contains a lot of unexplained variation.

5. Diagnosis

After the initial visualization, we can use diagnostic plotting to help in understanding the autocorrelation in the data.



The Autocorrelation Function (ACF) plot shows a gradual decline and significant autocorrelation at many lags. This suggests that there is a strong persistent pattern in the data, with past values having a long-lasting influence on future values. The slow decay of the ACF also indicates that the series is likely non-stationary, which is in line with the results of the ADF and KPSS tests previously discussed.



The Partial Autocorrelation Function (PACF) plot shows a significant spike at lag 1, after which the correlations quickly drop off and remain within the confidence interval (the blue shaded area). This suggests that there is only one significant autoregressive term when the effects of the other lags are accounted for, implying an AR(1) process may be a good starting point for modeling the data.

6. Damage

Outliers are associated with the damage that can skew the analysis, and if they correspond with significant market events, they could represent leverage points that need to be addressed.

7. Directions

Given the strong autocorrelation indicated by the ACF plot and the fact that the PACF plot suggests an AR(1) process, the data is non-stationary, the next steps in modeling the Apple stock price series could involve:

ARIMA Model: Based on the PACF plot, an initial model of ARIMA(1,1,0) could be considered, where '1' represents the AR term based on the PACF, '1' represents the order of differencing suggested by the ACF plot, and '0' for the MA term as there are no significant spikes in the ACF plot after lag 1.

Model Validation: After fitting the model, apply residuals to ensure data is white noise, indicating a good fit.

References

1. Enders, W. (2014). "Applied Econometric Time Series." Wiley.