Time series

Mohamed Imran

Data Scientist – Shell

What is a Time Series?

- Series of data points plotted over time.
- It is a sequence taken at successive equally spaced points in time.
- Used in forecasting.

How did our forefathers do the forecasting?

Traditional methods:

- **SMA** (*Simple Moving Average/Rolling Statistics*)
- **WMA** (*Weighted Moving Average*)
- **EWMA** (*Exponential Weighted Moving Average*)

SMA:

A simple moving average is calculated from the average of the closing prices for the time period being examined.

Date	Turnover
June	$250, 000.00
July	$901, 050.00
August	$501, 010.00
September	$700, 011.00
October	$900,100.00

The SMA(3) for the month of **November** is: (501,010 + 700,010 + 900,100)/3 = **700,373**

The SMA(3) for the month of **December** is: (700,010 + 900,100 + 700,373)/3 = **766,827**

WMA

Weighted moving average assigns a heavier weighting to more current data points since they are more relevant than data points in the distant past. The sum of the weighting should add up to 1 (or 100%).

Date	Closing Price	Weighting
October	$90.90	5/15
September	$90.36	4/15
August	$90.28	3/15
July	$90.83	2/15
June	$90.91	1/15

- The denominator is calculated as : 1+2+3+4+5 = 15
- In the example above, the weighted **5-day** moving average is $90.62
- ((90.9*(5/15))+(90.36*(4/15))+(90.28*(3/15))+(90.83*(2/15))+(90.91*(1/15)))

EWMA:

An exponential moving average (EMA) is similar to SMA, but whereas SMA removes the oldest prices as new prices become available, an exponential moving average calculates the average of all historical ranges, starting at the point you specify.

To calculate EMA, take current price and multiply it by a constant, C. Take previous period's EMA and multiply it by 1 minus that constant, C. Add the two values together.

$$EMA_{current} = C * (Price_{current}) + (1-C) * (EMA_{previous\ period})$$

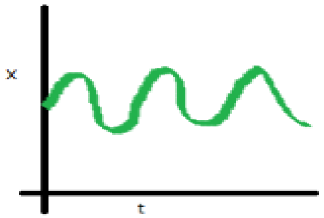
If you are calculating your first EMA value where there is no previous day's EMA, use SMA instead.
The formula for deriving the value of the constant, C is:

$$C = \frac{2}{(\#\ of\ periods + 1)}$$

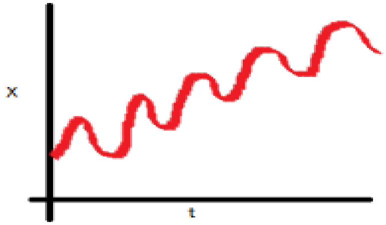
What is Stationary?

What does it mean for data to be stationary?

1. The mean of the series should not be a function of time. The red graph below is not stationary because the mean increases over time.

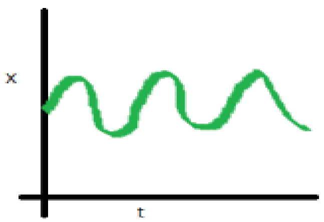


Stationary series

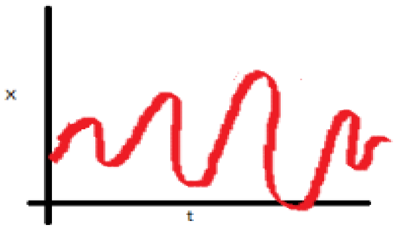


Non-Stationary series

2. The variance of the series should not be a function of time. This property is known as homoscedasticity. Notice in the red graph the varying spread of data over time.

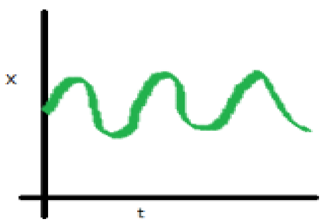


Stationary series

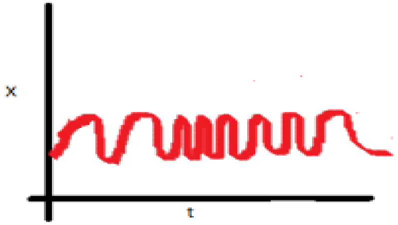


Non-Stationary series

3. Finally, the covariance of the i th term and the (i + m) th term should not be a function of time. In the following graph, you will notice the spread becomes closer as the time increases. Hence, the covariance is not constant with time for the 'red series'.



Stationary series



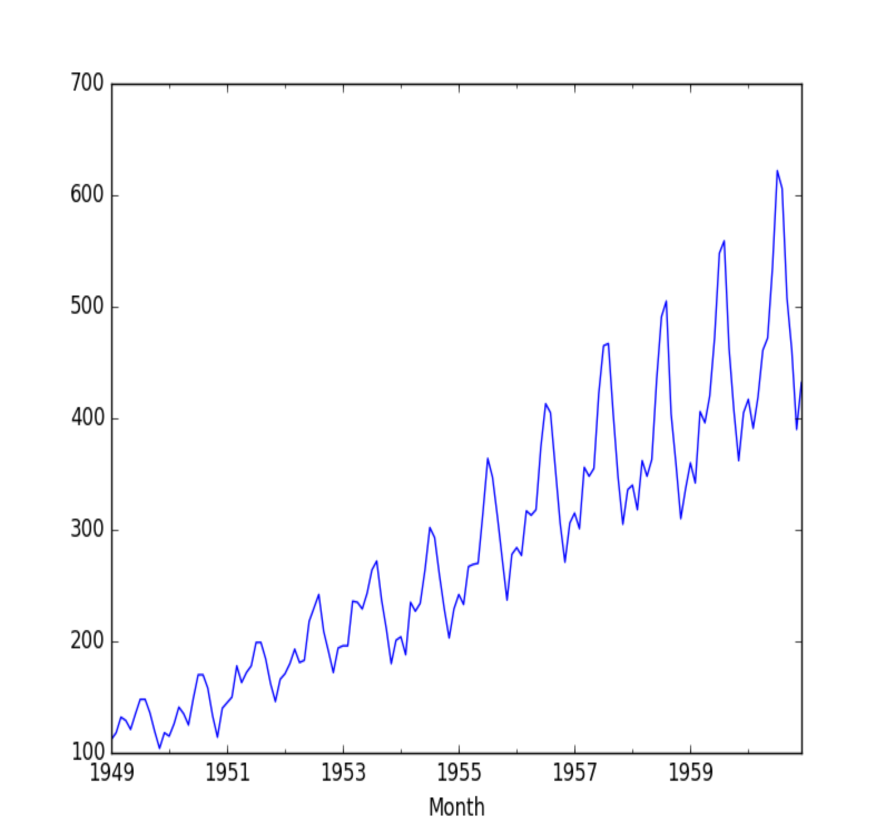
Non-Stationary series

Checks for Stationarity:

- No mean
- No Variance/Covariance

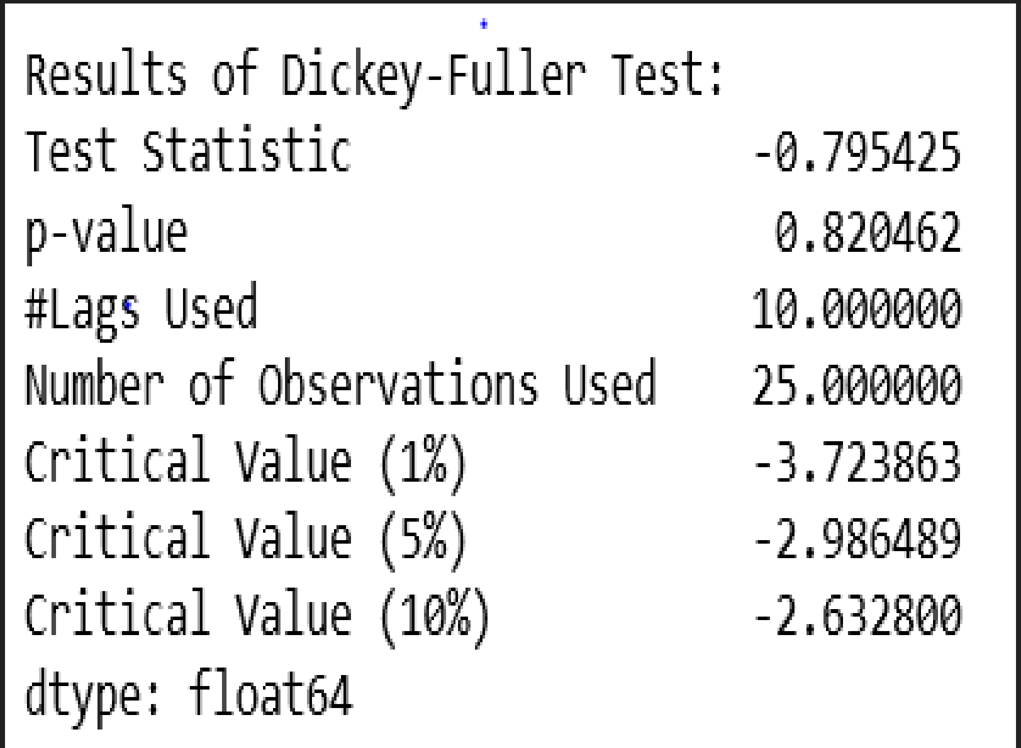
Rolling mean:

- Take the past year's data (12 months) and calculate the mean.
- Compare it with the successive years.
- If mean is a function of time (mean increases), the Time Series in not Stationary.



Dickey-Fuller test:

- If the test statistics is greater than the critical values, then the series is not Stationary.



Results of Dickey-Fuller Test:

Test Statistic	-0.795425
p-value	0.820462
#Lags Used	10.000000
Number of Observations Used	25.000000
Critical Value (1%)	-3.723863
Critical Value (5%)	-2.986489
Critical Value (10%)	-2.632800

dtype: float64

So, what is making a Time Series non-stationary?

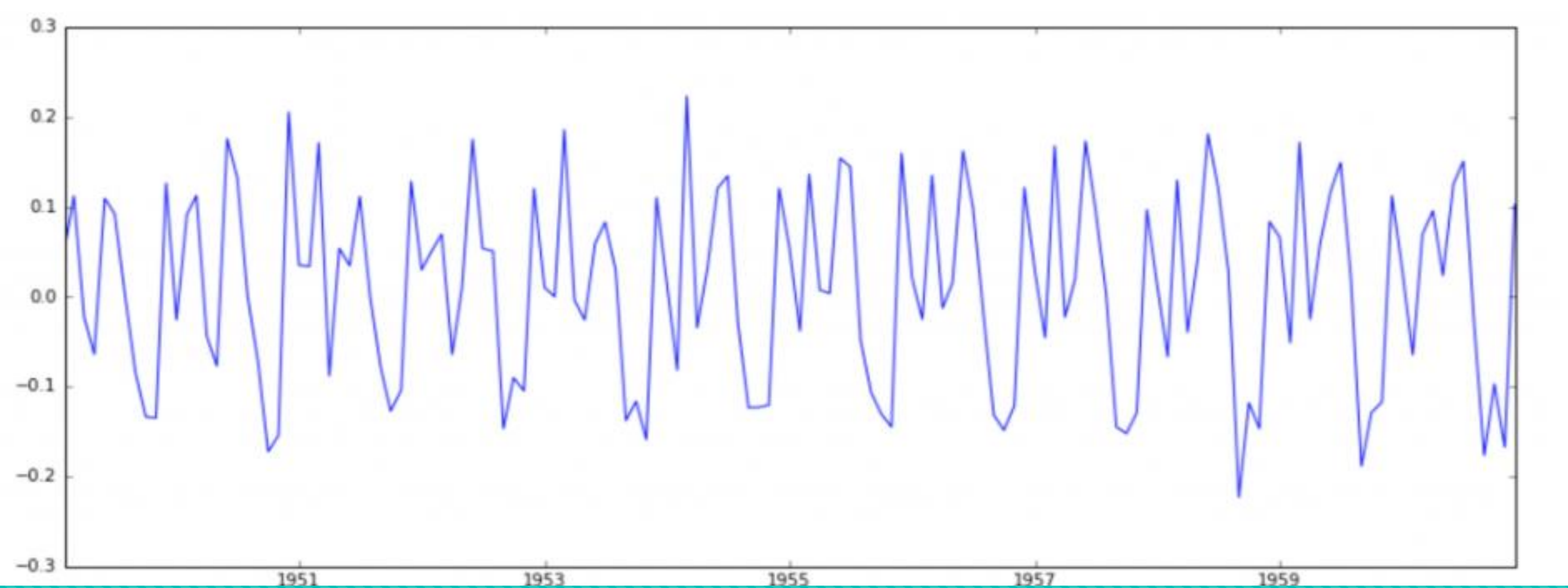
So, what is making a Time Series non-stationary?

- Trend
- Seasonality

Well, how do we remove Trend and Seasonality?

Well, how do we remove Trend and Seasonality?

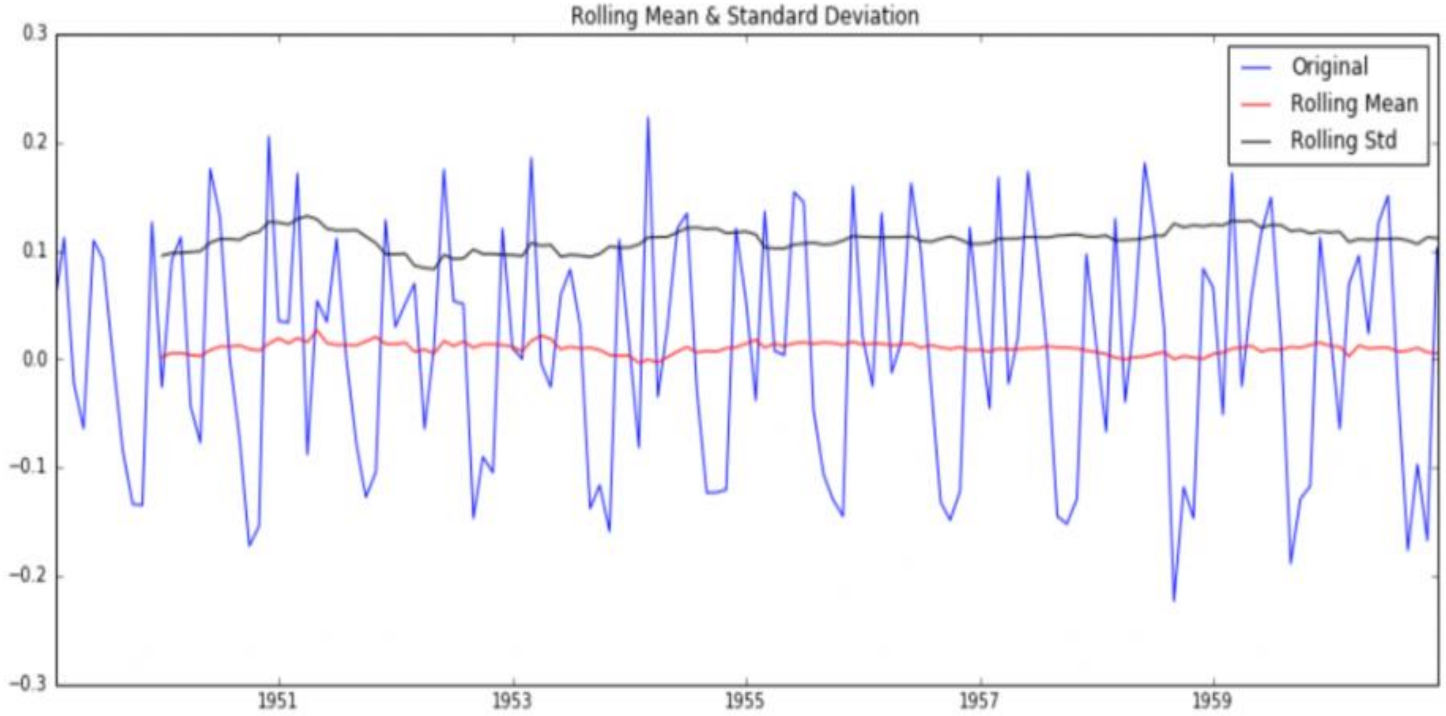
- Differencing
- Decomposition



Differencing:

Taking difference of an observation of a particular instance with that of a previous instance.

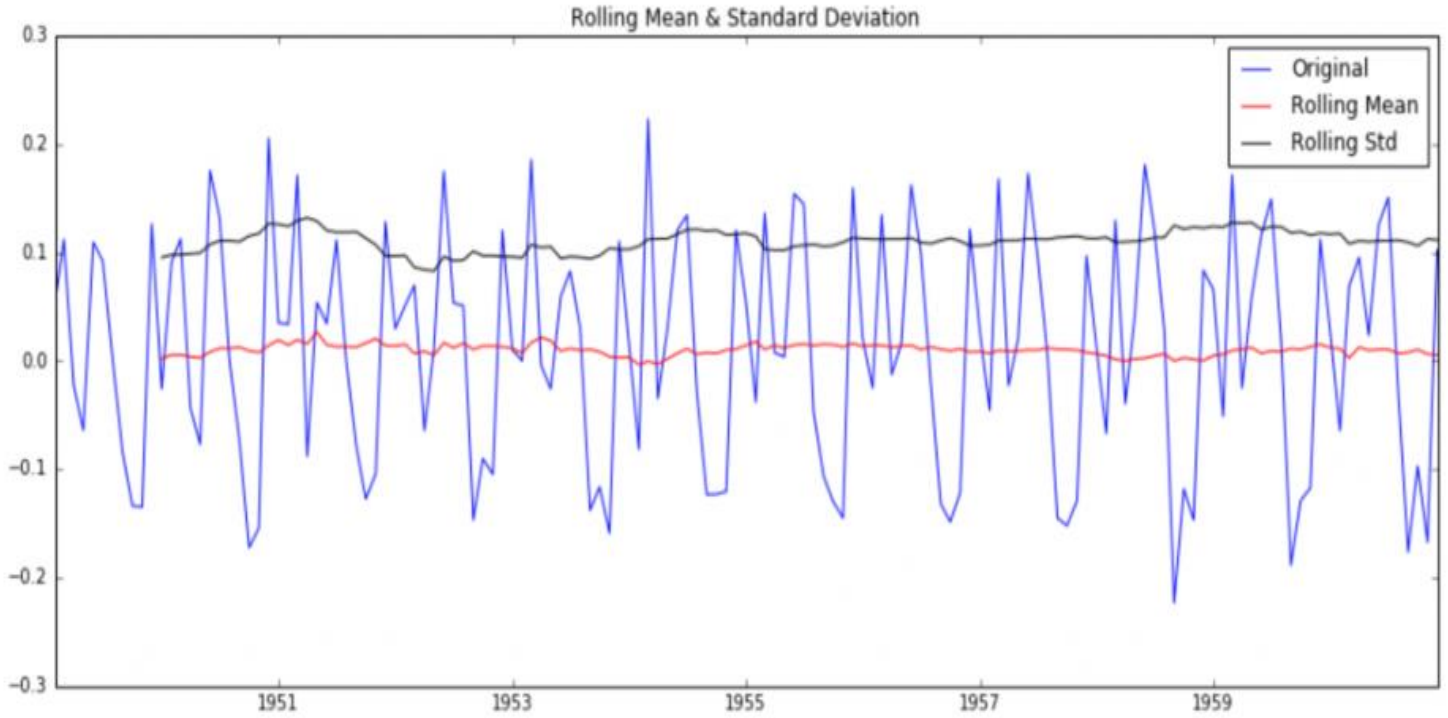
What are the observations as a result of Differencing?



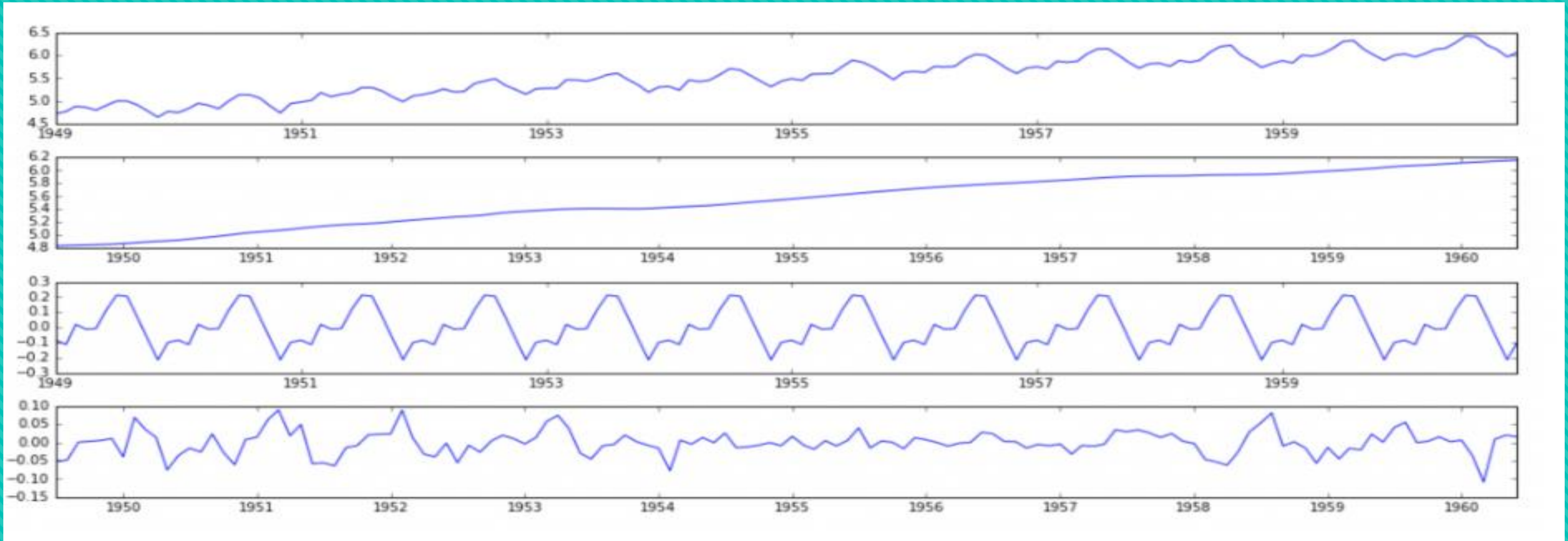

```
Results of Dickey-Fuller Test:
Test Statistic                  -2.717131
p-value                          0.071121
#Lags Used                      14.000000
Number of Observations Used    128.000000
Critical Value (5%)             -2.884398
Critical Value (1%)             -3.482501
Critical Value (10%)            -2.578960
dtype: float64
```

Rolling mean and standard deviation is constant.

Dickey-Fuller Test Statistics is less than **10%** of critical values

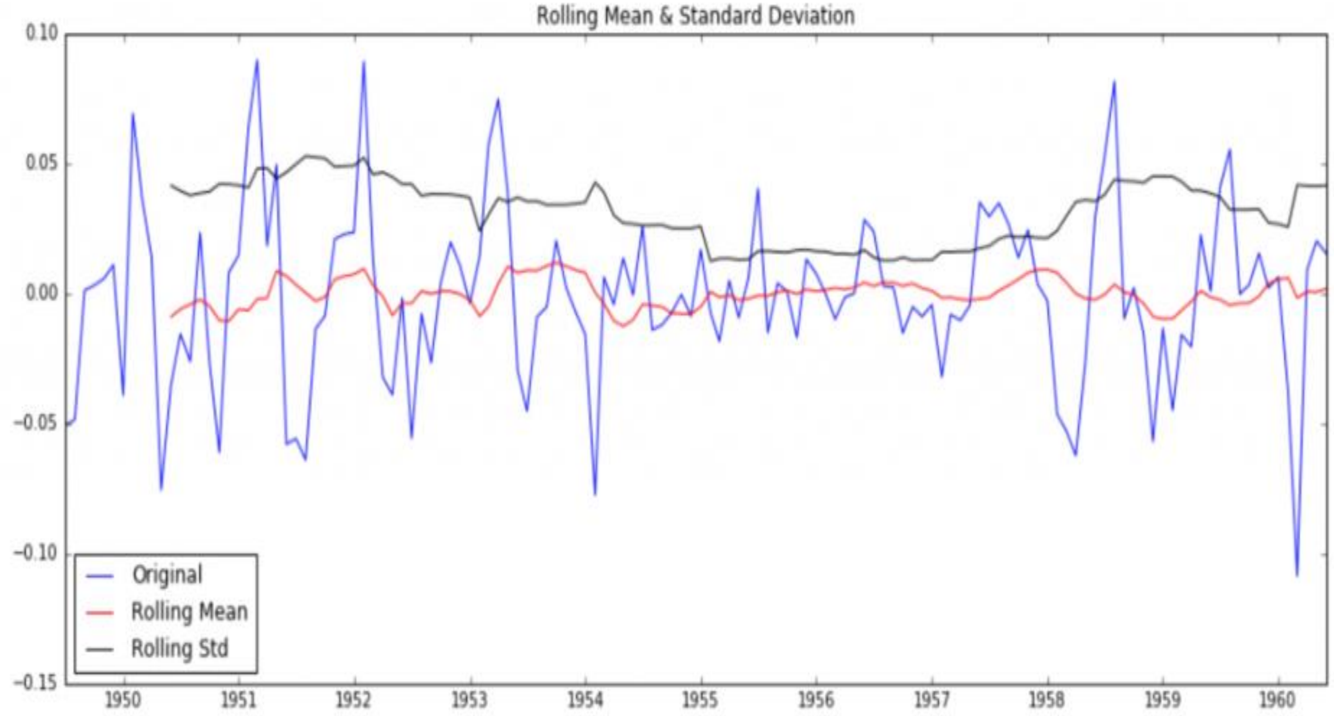



```
Results of Dickey-Fuller Test:
Test Statistic                  -2.717131
p-value                          0.071121
#Lags Used                      14.000000
Number of Observations Used    128.000000
Critical Value (5%)             -2.884398
Critical Value (1%)             -3.482501
Critical Value (10%)            -2.578960
dtype: float64
```



Decomposition:

The trend and seasonality are modeled separately, then the rest of the series is returned




```
Results of Dickey-Fuller Test:
Test Statistic                  -6.332387e+00
p-value                          2.885059e-08
#Lags Used                       9.000000e+00
Number of Observations Used      1.220000e+02
Critical Value (5%)             -2.885538e+00
Critical Value (1%)             -3.485122e+00
Critical Value (10%)            -2.579569e+00
dtype: float64
```

Mean and standard deviation has improved

The Dickey-Fuller test statistic is significantly **lower than the 1% critical value**. So this TS is very close to stationary.

Forecasting

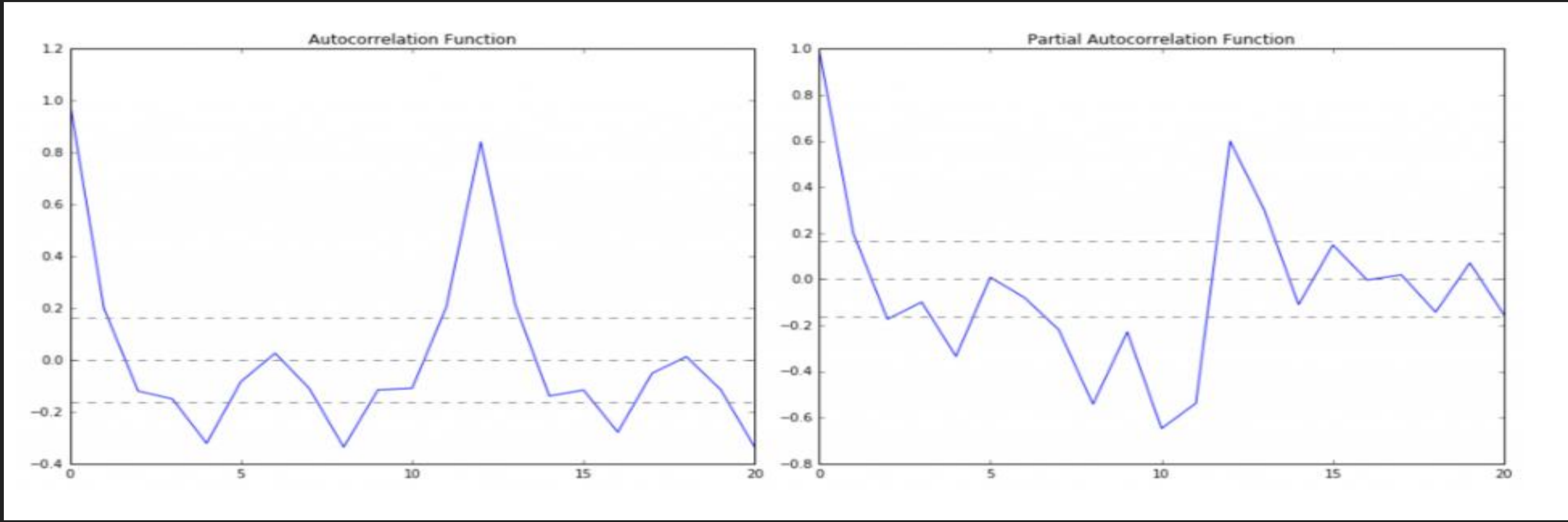
ARIMA:

- Auto-Regressive Integrated Moving Averages.
- It depends on the parameters (p, d, q)
- p - is just lags of dependent variable.
- q – is lagged forecast errors in prediction equation.
- d - the number of non-seasonal differences

How to determine p and q values?

p – The lag value where the **PACF** chart crosses the upper confidence interval for the first time.

q – The lag value where the **ACF** chart crosses the upper confidence interval for the first time.



Auto-Arima

- Auto arima gives the p, d and q values.
- Run the function with the Time Series data to get these values.

Other forecasting algorithms in R:

- HW
- BSTS
- Croston
- Arima_NS
- Arima_NS_Boxcox
- fbprophet