Modeling and Forecasting with ARMA Processes

Agenda

- Lag and Differencing
- ACF and PACF
- What is stationary Processes ?
- How to find Stationarity ?
- ARMA(p, q) Processes

Lag values in Time Series

Lag

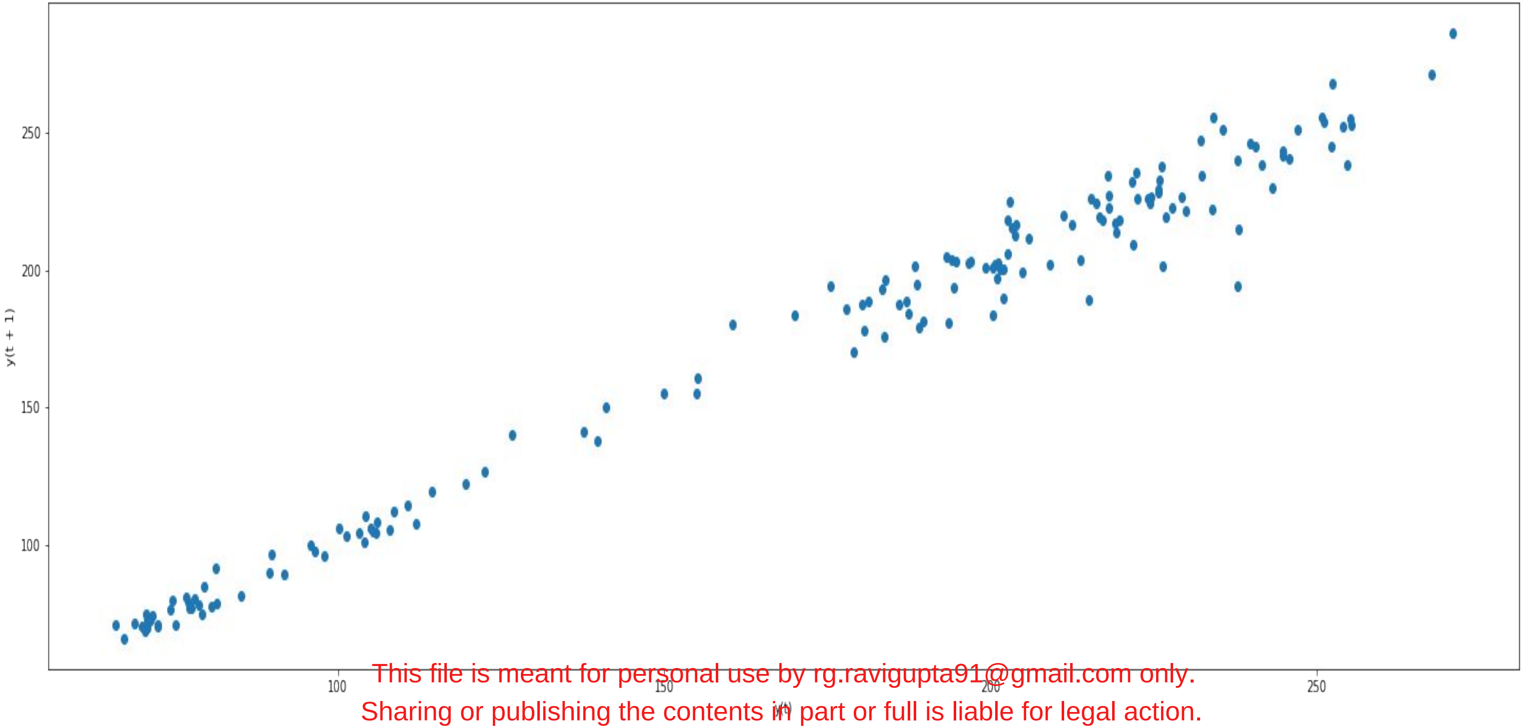
- In time series, effect of many variable can last beyond it happens. Example advertisement campaign.
- Therefore lagged values can be useful for forecasting the time series.
- Lag is yesterday's observation today

BASF stock price TS

Year	Month	Original Series	Lag(1) Series	Lag(2) Series	Lag(3) Series
1981	Jan	67.27			
1981	Feb	65.86	67.27		
1981	Mar	70.80	65.86	67.27	
1981	Apr	72.38	70.80	65.86	67.27
1981	May	70.61	72.38	70.80	65.86
1981	Jun	74.62	70.61	72.38	70.80
1981	Jul	79.56	74.62	70.61	72.38
1981	Aug	85.08	79.56	74.62	70.61
1981	Sep	81.39	85.08	79.56	74.62
1981	Oct	78.73	81.39	85.08	79.56
1981	Nov	78.01	78.73	81.39	85.08

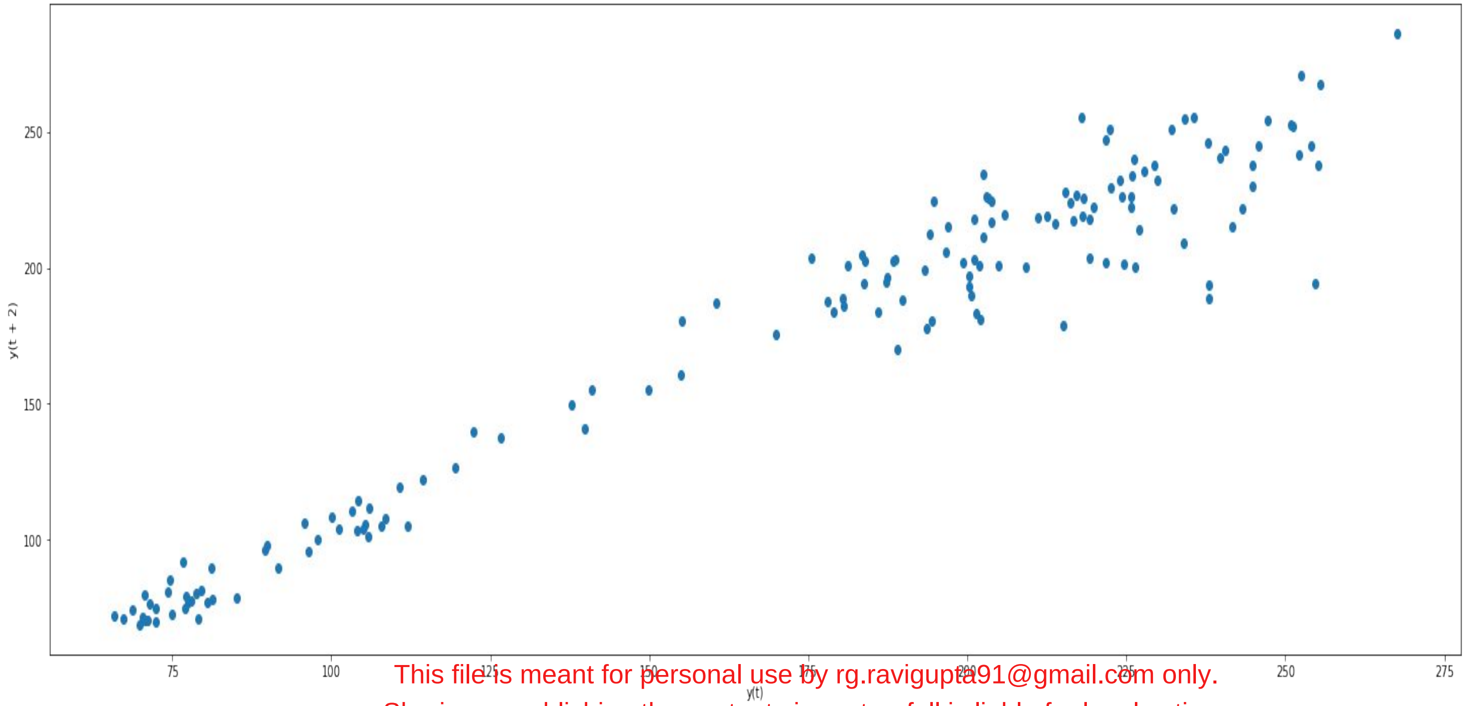
Lag plot of y_t and $y_{(t+1)}$

BASF stock price TS



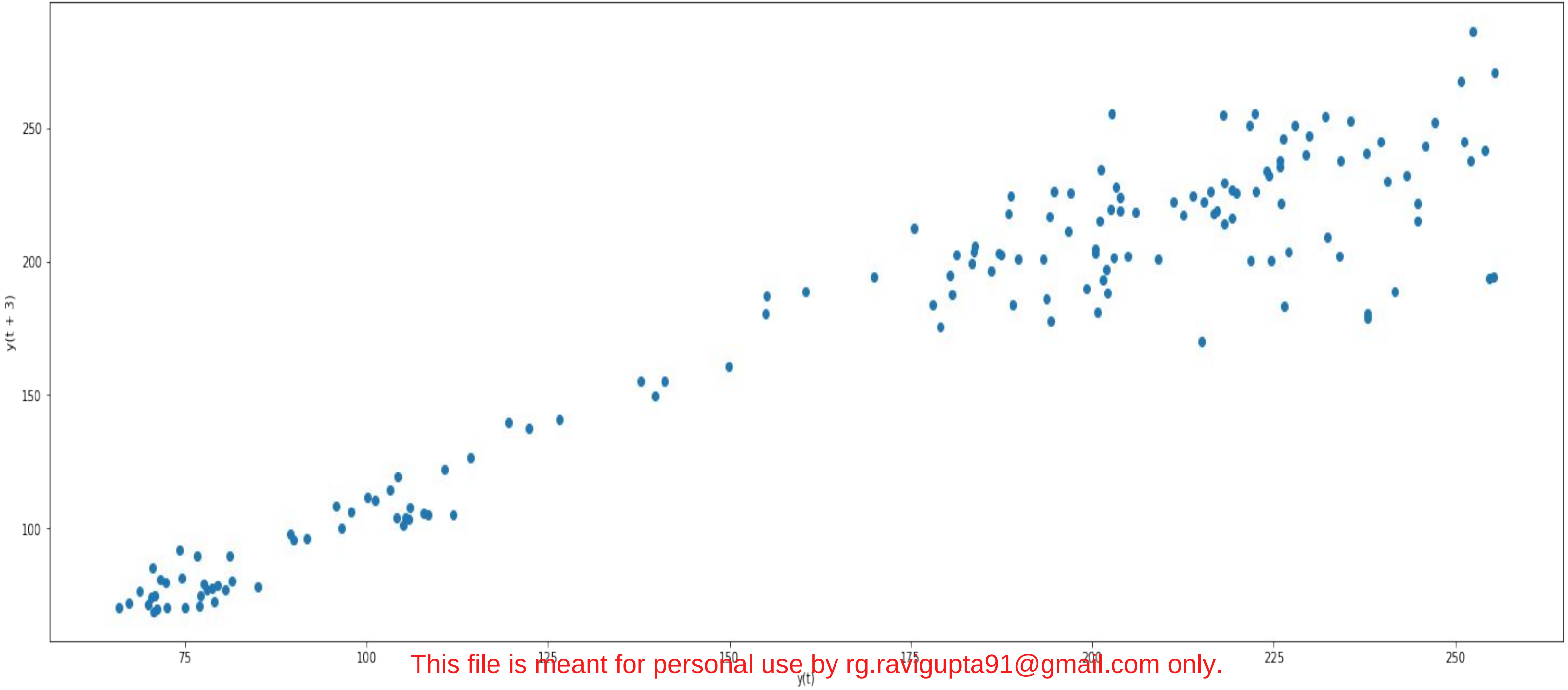
Lag plot of y_t and $y_{(t+2)}$

BASF stock price TS



Lag plot of y_t and $y_{(t+3)}$

BASF stock price TS



ACF and PACF

Autocorrelation Function

- An Autocorrelation function (ACF) determines the average correlation between time series observations and its past values for different lag.

- Example, the correlation at lag 1 is the correlation between observations of the time series measured at time t with all the observations at time period t − 1.

- The correlation at lag 2 is the correlation between observations of the time series measured at time t with all of the observations at time period t − 2.

Partial Autocorrelation Function for ARMA process

- A partial autocorrelation function is the correlation between time series observation and a lag of itself that is not explained by correlations at all lower-order-lags.

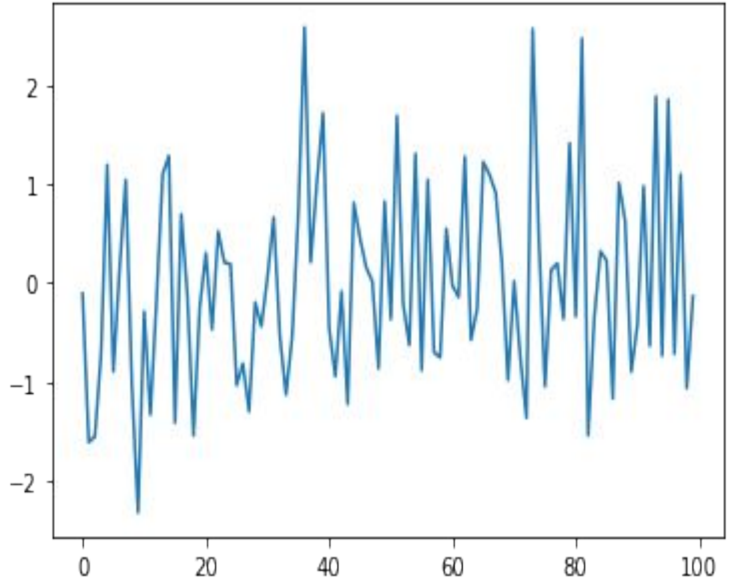
- A partial autocorrelation function is similar to an autocorrelation function except that each correlation controls for any correlation between observations of a shorter lag length.

Partial Autocorrelation Function for ARMA process

- Thus, the value for the ACF and the PACF at the first lag are the same because both measure the correlation between data points at time t with data points at time $t-1$.

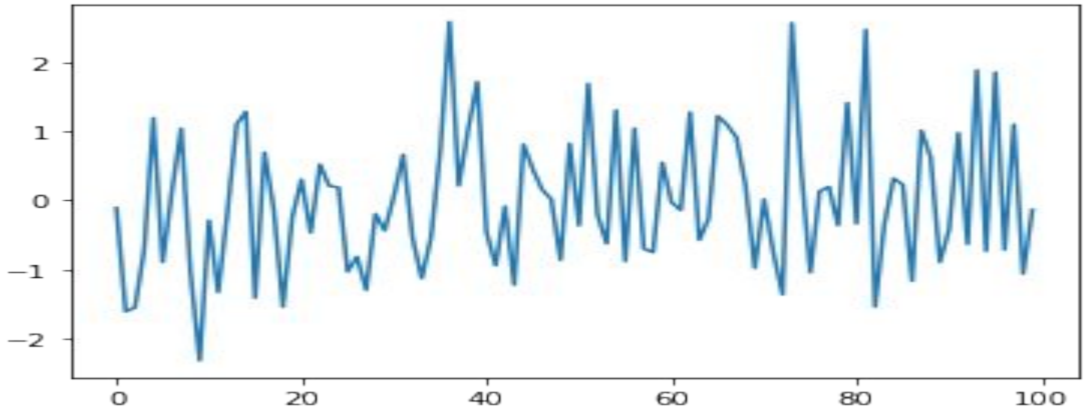
- However, at the second lag, the PACF measures the correlation between data points at time t with data points at time $t-2$ after controlling for the correlation between data points at time t with those at time $t-1$.

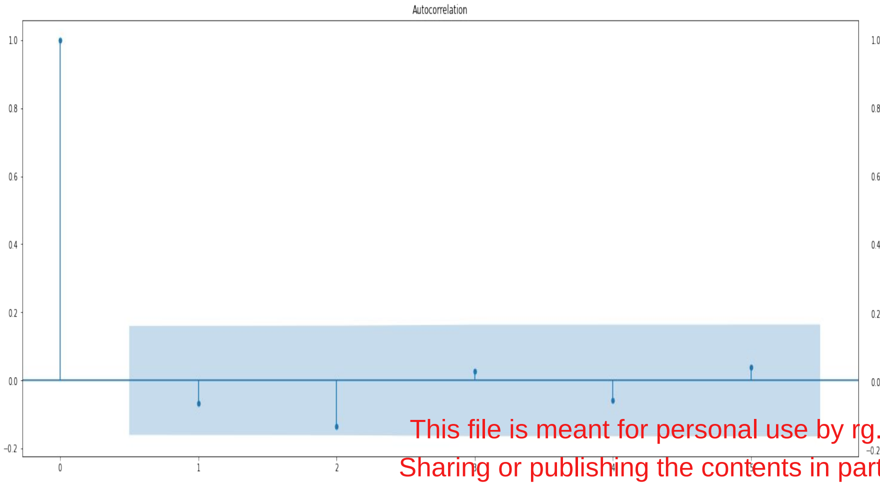
For random process with μ= 0 and σ= 1

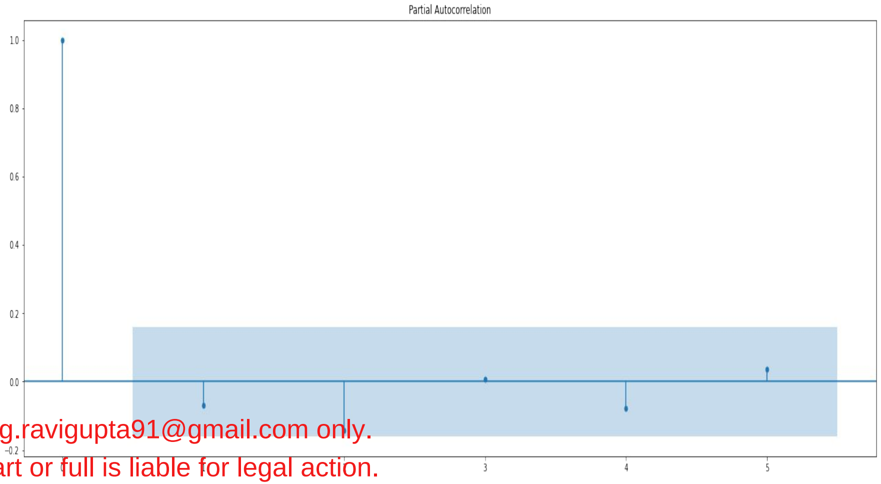


Lag	ACF	PACF
0	1	1
1	-0.067	-0.068
2	-0.135	-0.143
3	0.027	0.007
4	-0.058	-0.079

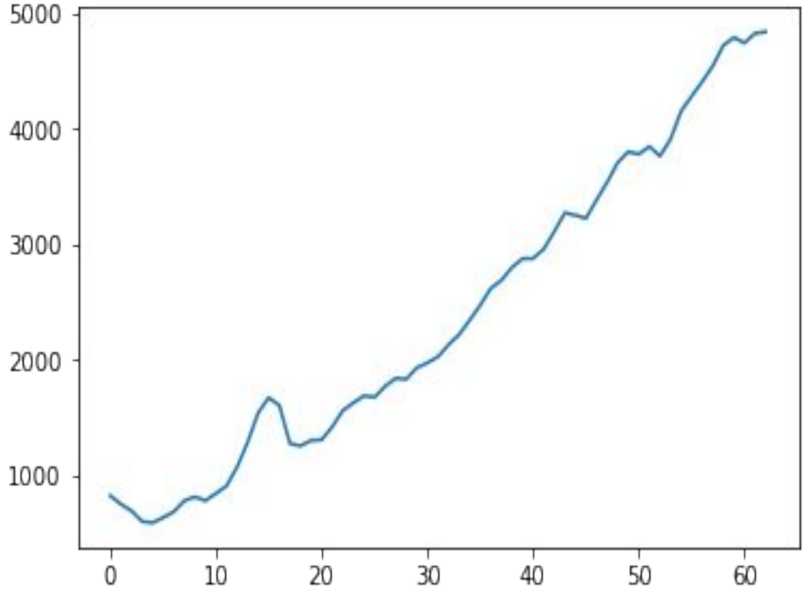
For random process with μ= 0 and σ= 1





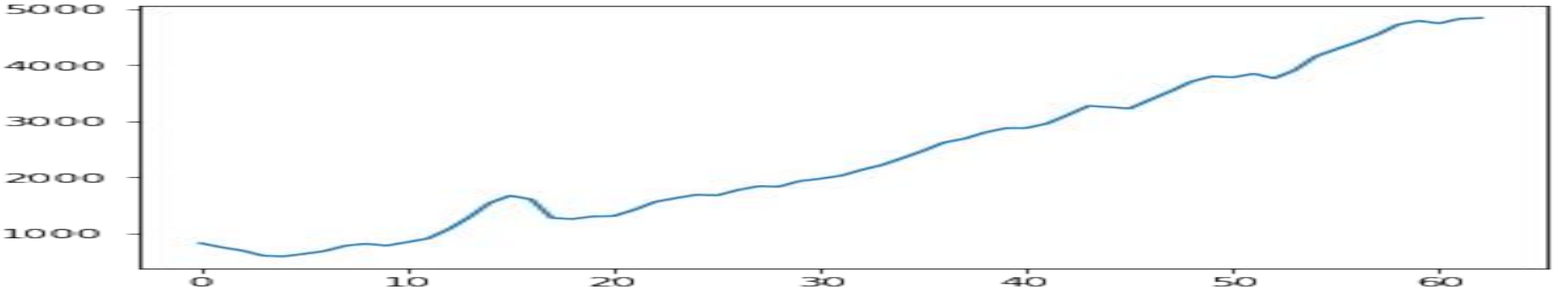


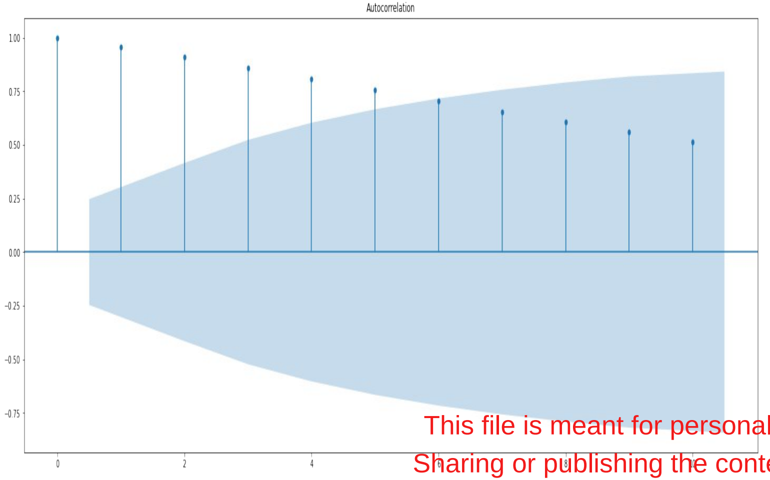
For series with trend

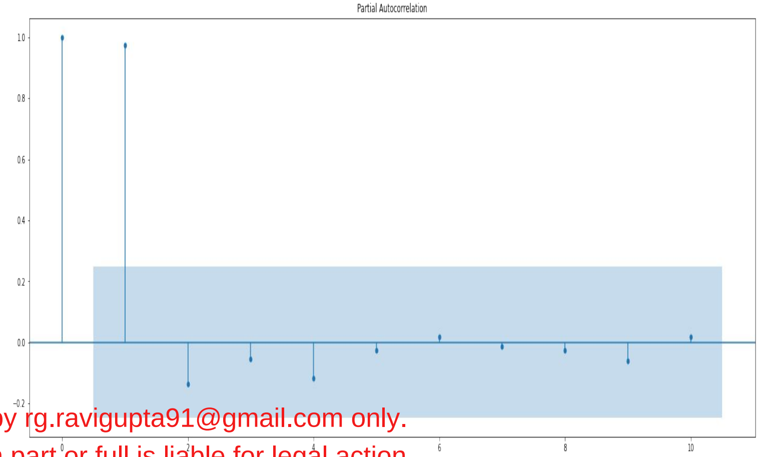


Lag	ACF	PACF
0	1	1
1	0.95	0.97
2	0.90	-0.13
3	0.86	-0.05
4	0.80	-0.11

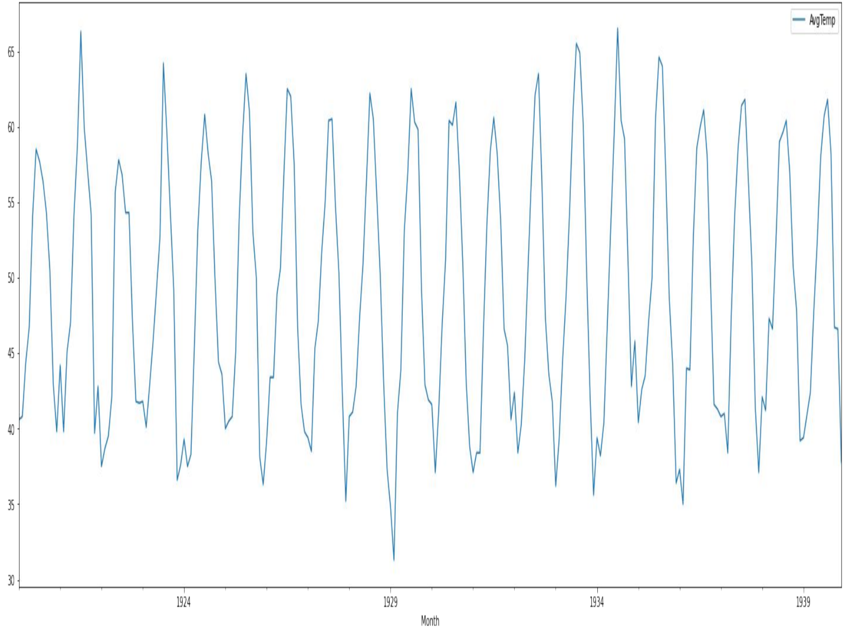
For series with trend





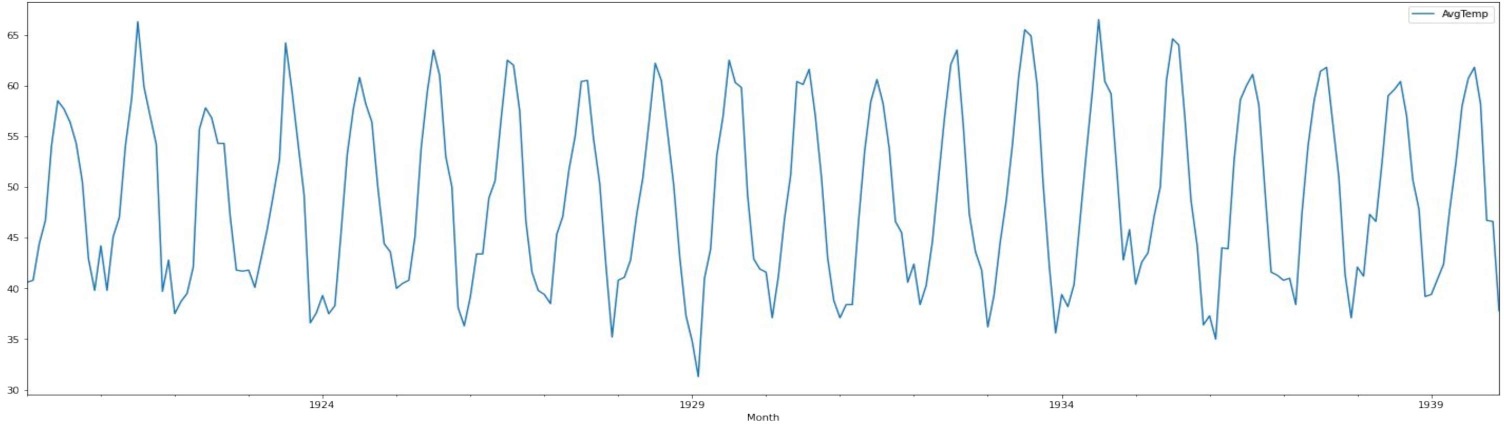


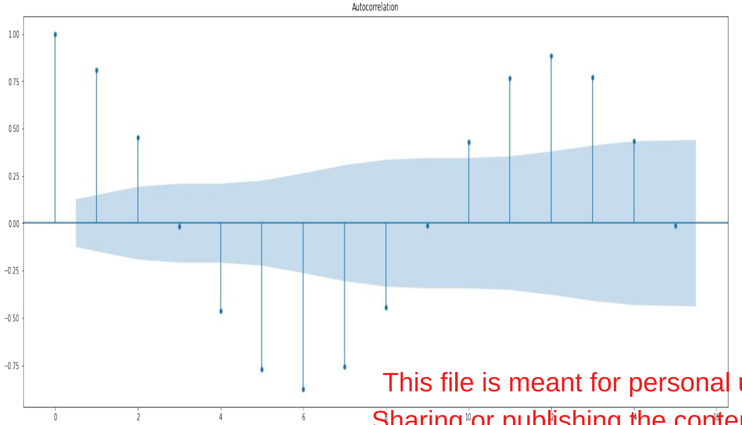
For series with seasonality

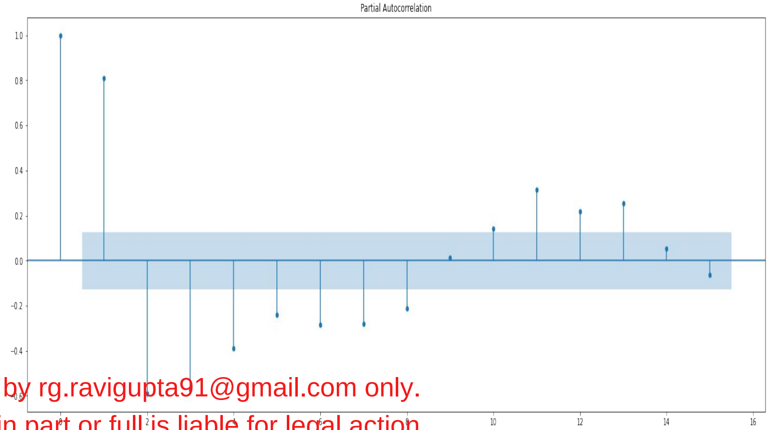


Lag	ACF	PACF
0	1	1
1	0.80	0.81
2	0.45	-0.59
3	-0.01	-0.57
4	0.46	-0.38

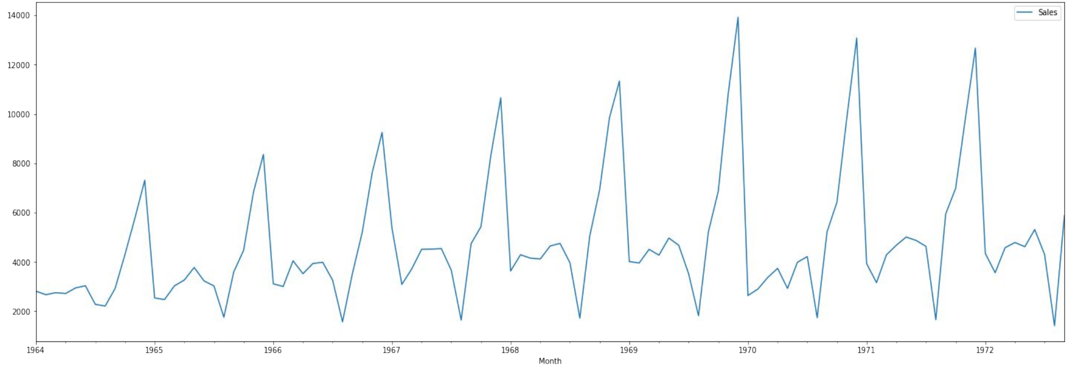
For series with seasonality

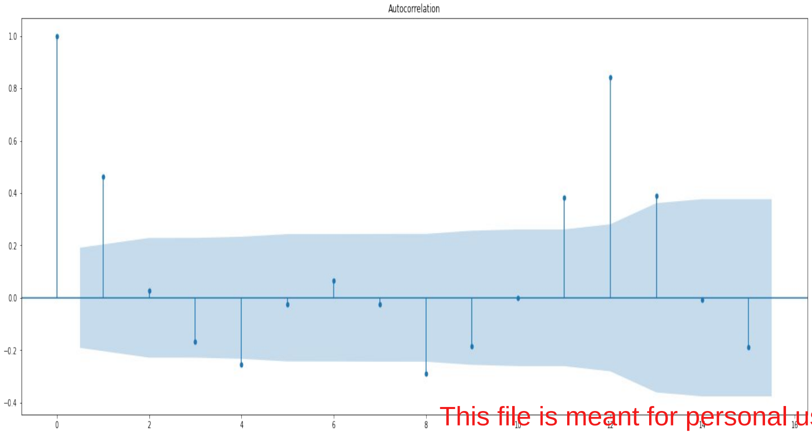


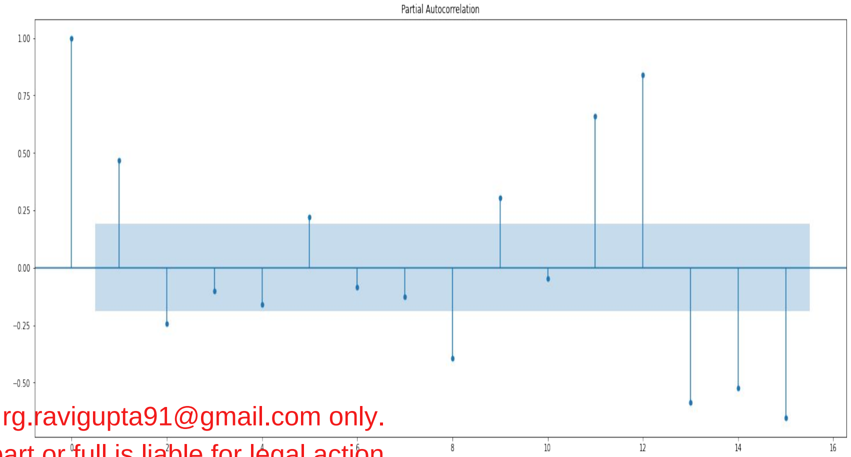




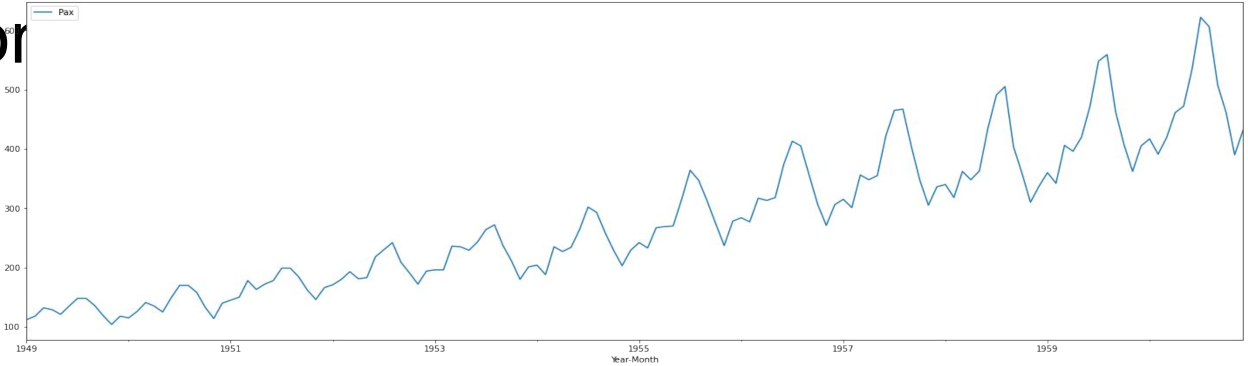
For series with trend and seasonality

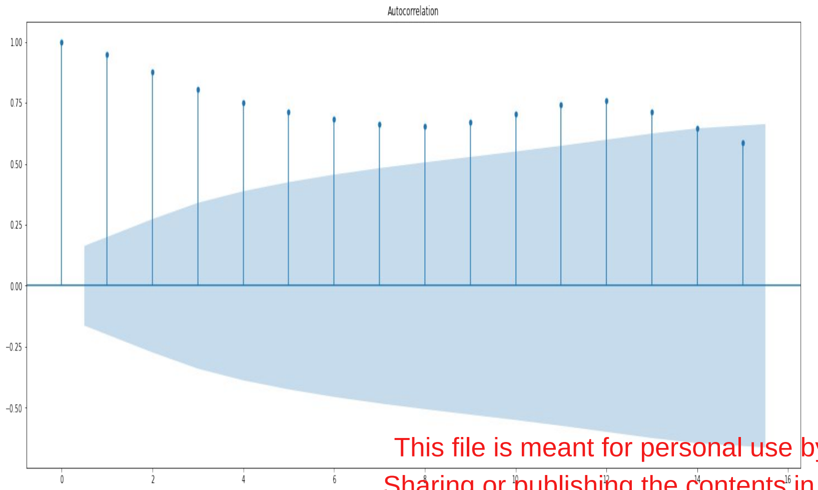


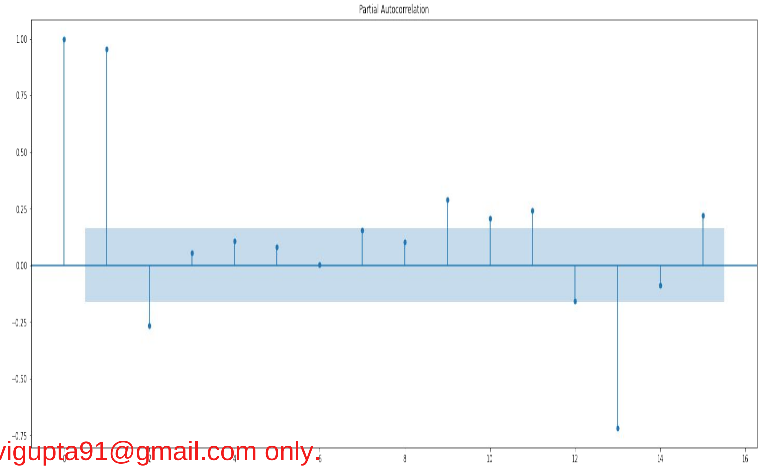




For series with trend and multiplicative season







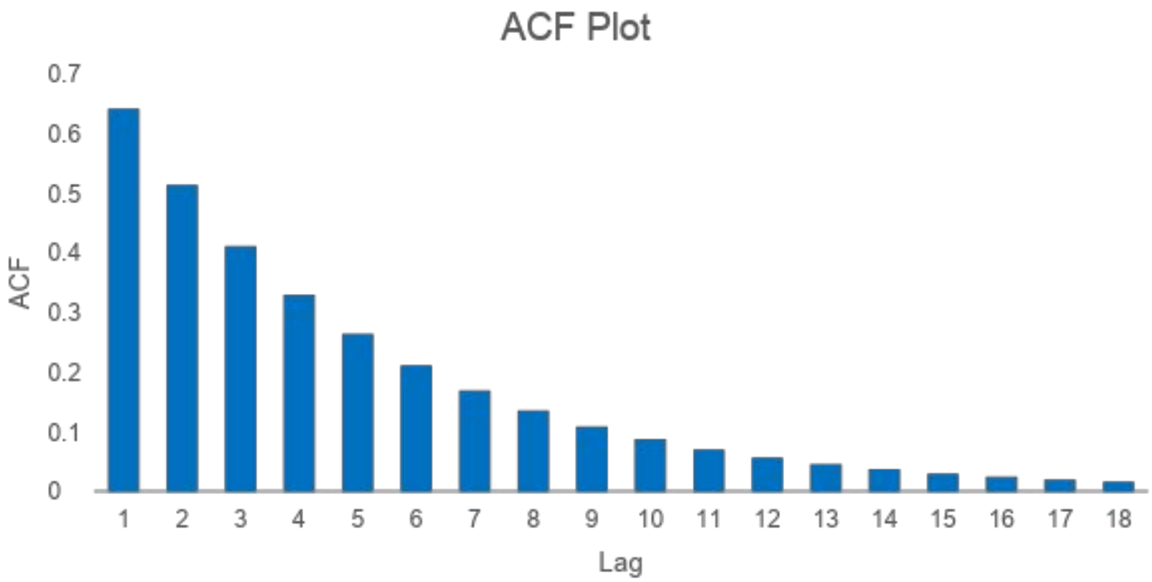
Understanding ACF

- Consider a time series having average correlation between observations recorded at time t with observations recorded at time t − 1 is 0.8 and there is no other pattern of correlation.
- The ACF value at the first lag would equal 0.8 but the ACF value at the second lag would be 0.64.
- This is because time series observations at time t are correlated with observations at time t−1 at 0.8 and observations at time t−1 are correlated with observations at time t−2 at 0.8.
- Therefore, observations at time t are correlated with observations at time t−2 at the level of $0.8 \times 0.8 = 0.64$.
- The ACF would continue to decline toward zero as the lag length increased

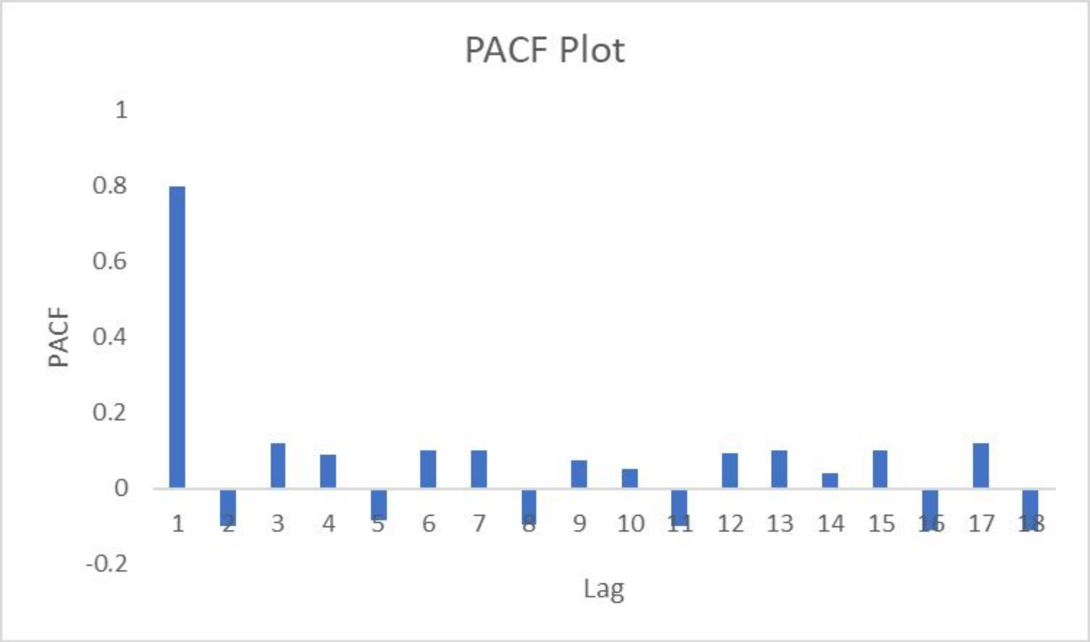
Understanding PACF

- For the same time series, the value of the PACF at the first lag would also be equal to 0.8.
- However, the value of the PACF at the second lag would be equal to zero, plus or minus some random error.
- This is because there would be no correlation between data points at time t and data point at time t − 2 after accounting for the fact that they are both correlated with data points at time t − 1.

ACF Plot

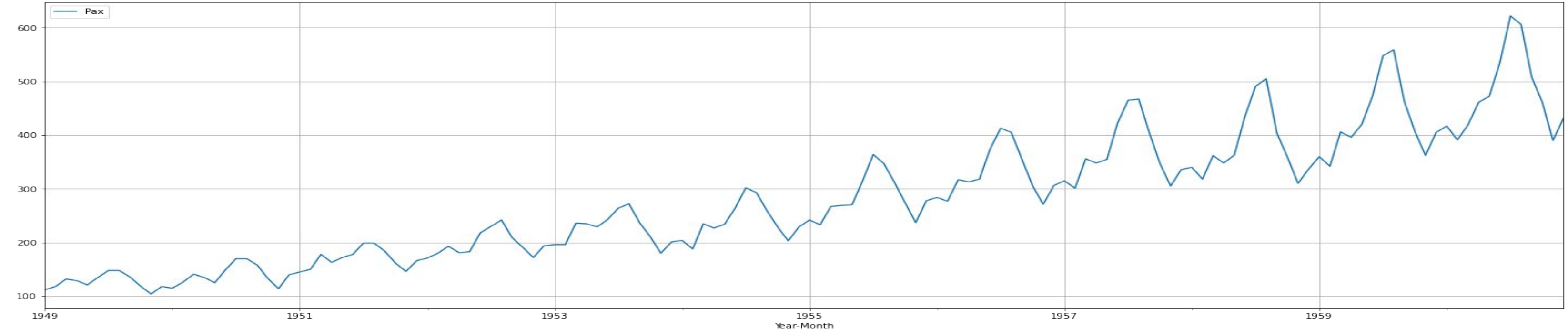

ACF Plot
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
ACF
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
Lag

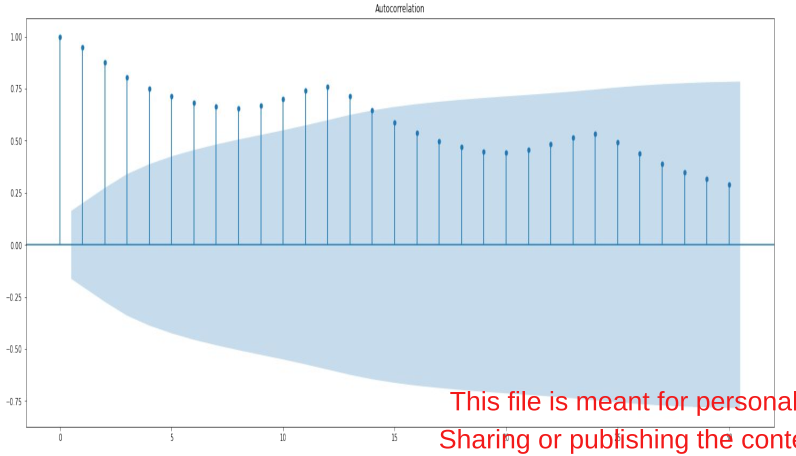
PACF plot

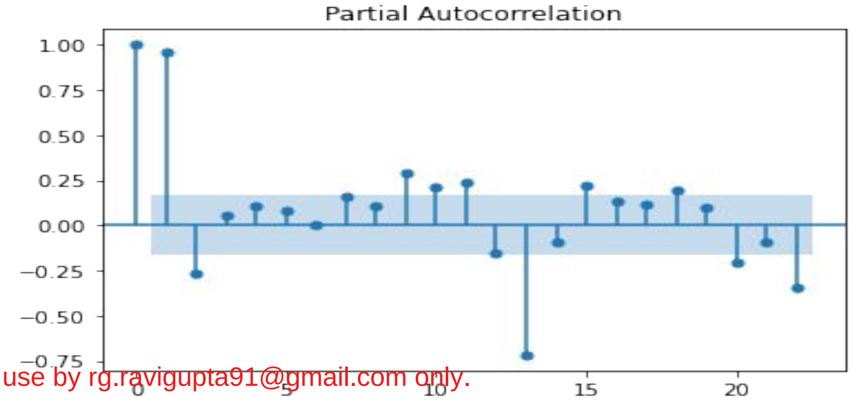


PACF Plot
1
0.8
0.6
0.4
0.2
0
-0.2
PACF
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18
Lag

ACF –PACF plots: Air Passenger TS







Stationarity

Stationary time series

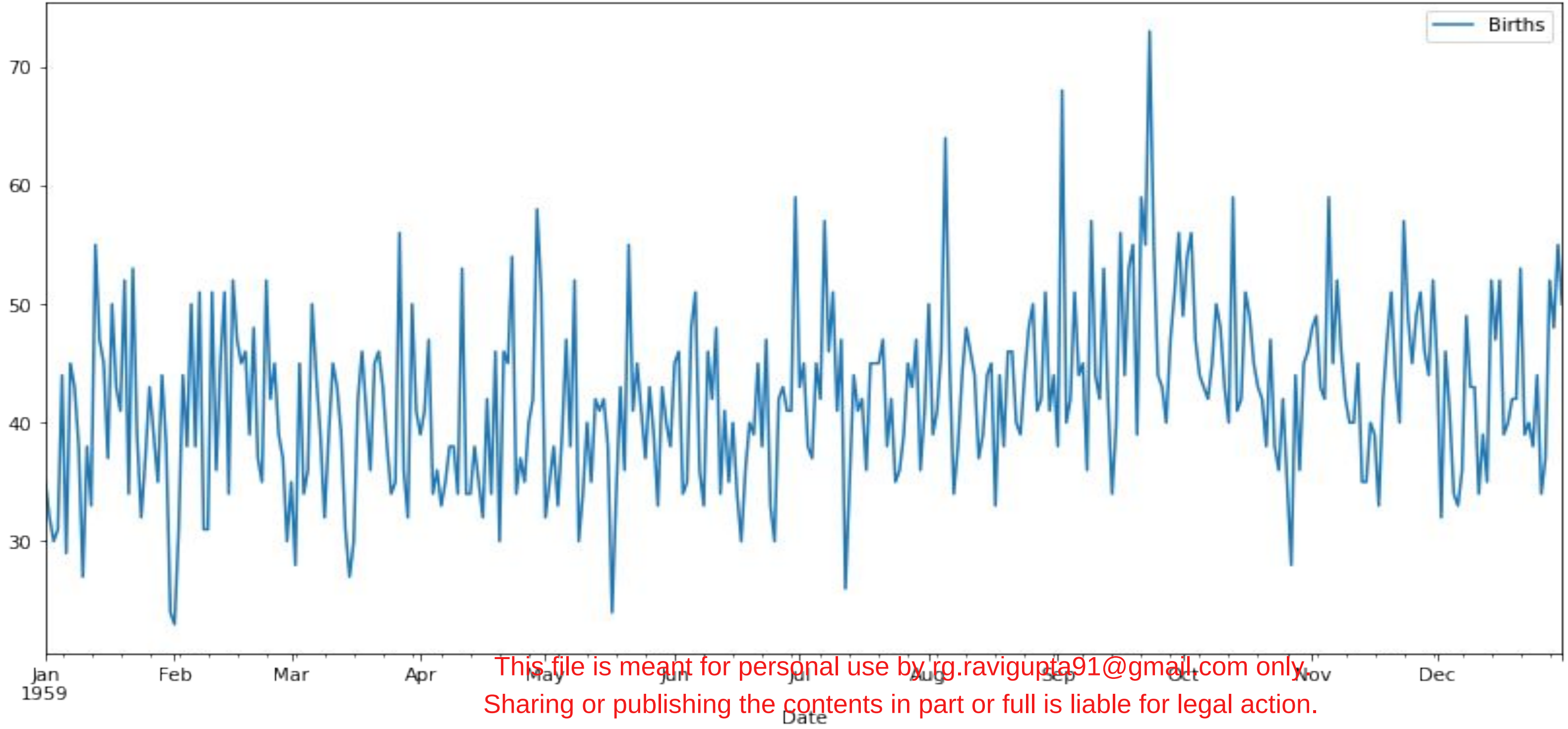
- Time series is said to be stationary, when its statistical properties does not change with respect to time.

- Stationary implies statistical equilibrium or stability in the data.

- Time series components like trend affects the values of time series at different time steps and therefore make the series non-stationary.

- In general, stationary time series plot is roughly horizontal without any predictable pattern in long term.

Stationary time series:- Daily total female birth data



Properties of stationary time series

- Stationary time series does not have pronounced trend.
- The mean and variance does not change with respect to time.
- Correlation depends on the lag only.

How to find Stationarity ?

- There are many methods to check whether a time series is stationary or not.
- Time series plot:
  - Time series can be checked for stationarity by visually inspecting the time series plot.
- Statistical test:
  - Statistical tests can be used to check the stationarity of time series.

Dicky Fuller test

- Dicky Fuller is the statistical test used for determining stationarity of time series.

- For D-F test,
  - null hypothesis: the time series is non-stationary.
  - alternate hypothesis: the time series is stationary.

- Test results can be interpreted using p-value.

Dicky Fuller test on female birth data

```
In [18]: df= pd.read_csv('daily-total-female-births.csv', header=0, index_col=0, squeeze=True)
```

```
In [19]: observations= df.values
         test_result = adfuller(observations)
```

```
In [20]: print('ADF Statistic: %f' % test_result[0])
         print('p-value: %f' % test_result[1])
         print('Critical Values:')
         for key, value in test_result[4].items():
             print('\t%s: %.5f' % (key, value))
```

```
ADF Statistic: -4.808291
p-value: 0.000052
Critical Values:
        1%: -3.44875
        5%: -2.86965
        10%: -2.57109
```

Differencing

- Differencing is the process of transforming a time series to stationary series.

- Used to replace the series with the difference between their current values and the previous values

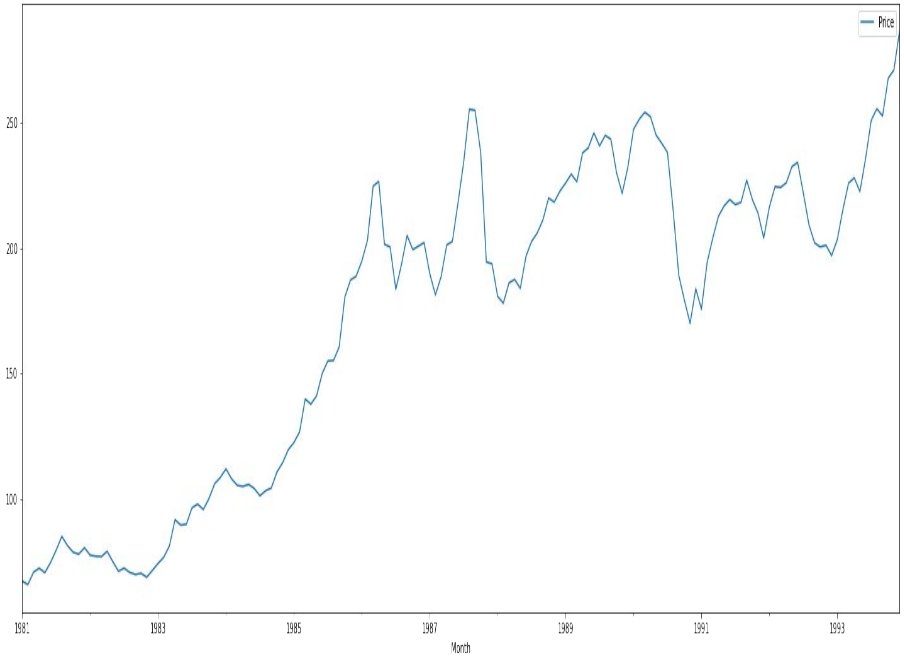
$$y_t = y_t - y_{t-1}$$

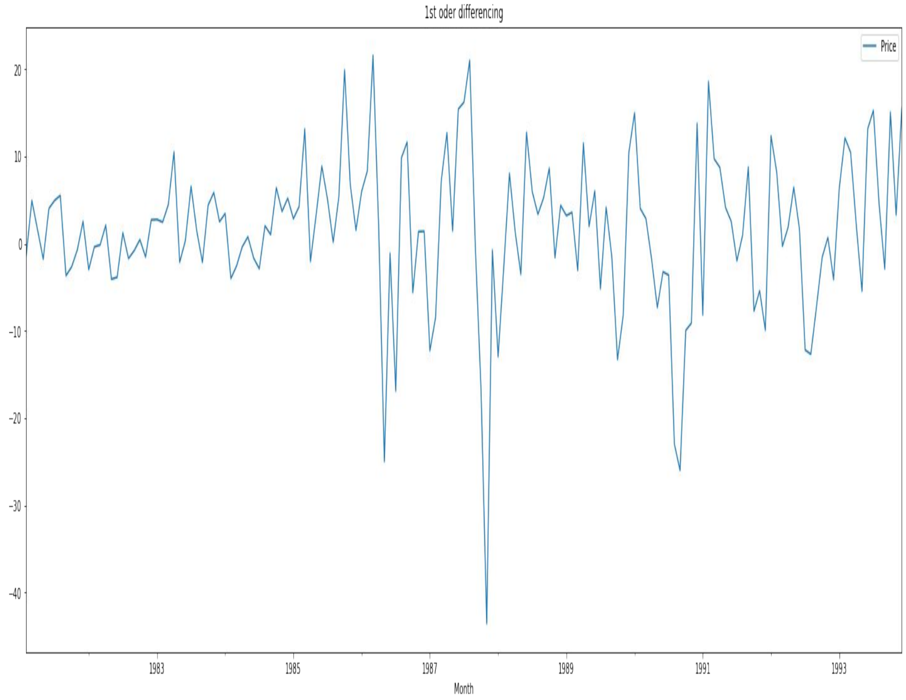
- Seasonal differencing removes the seasonality from the series.

BASF stock price TS

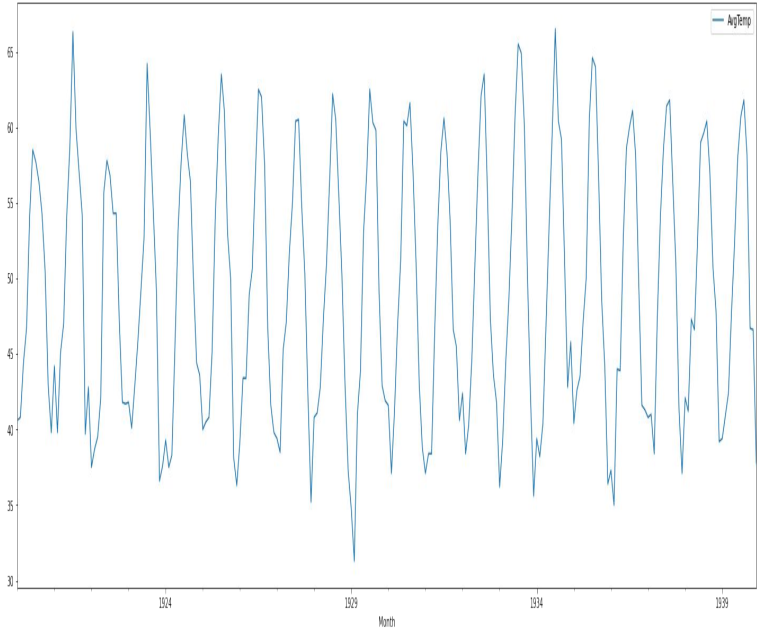
Year	Month	Original Series	Difference Series
1981	Jan	67.27	
1981	Feb	65.86	-1.40
1981	Mar	70.80	4.94
1981	Apr	72.38	1.57
1981	May	70.61	-1.77
1981	Jun	74.62	4.02
1981	Jul	79.56	4.94
1981	Aug	85.08	5.52
1981	Sep	81.39	-3.69
1981	Oct	78.73	-2.67
1981	Nov	78.01	-0.72

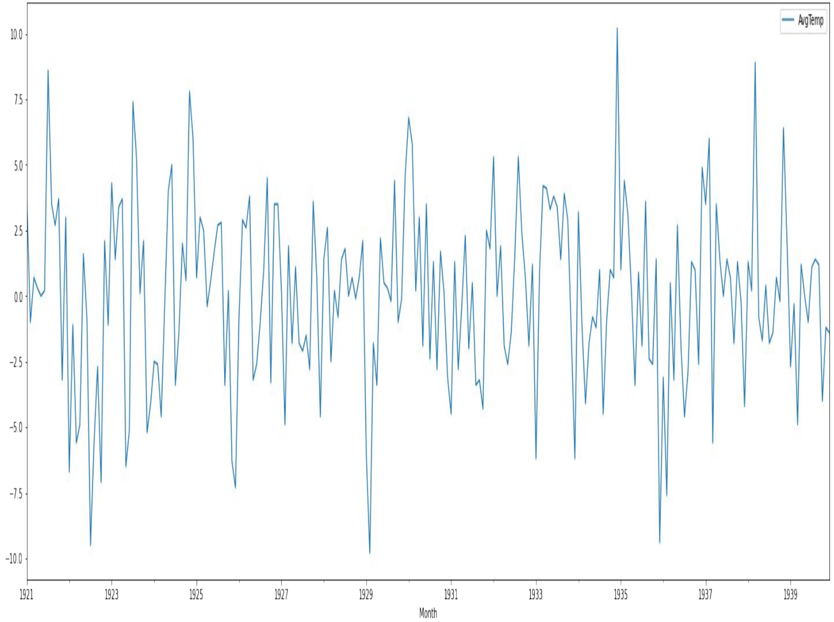
BASF stock price TS: 1^{st} order differencing for trend



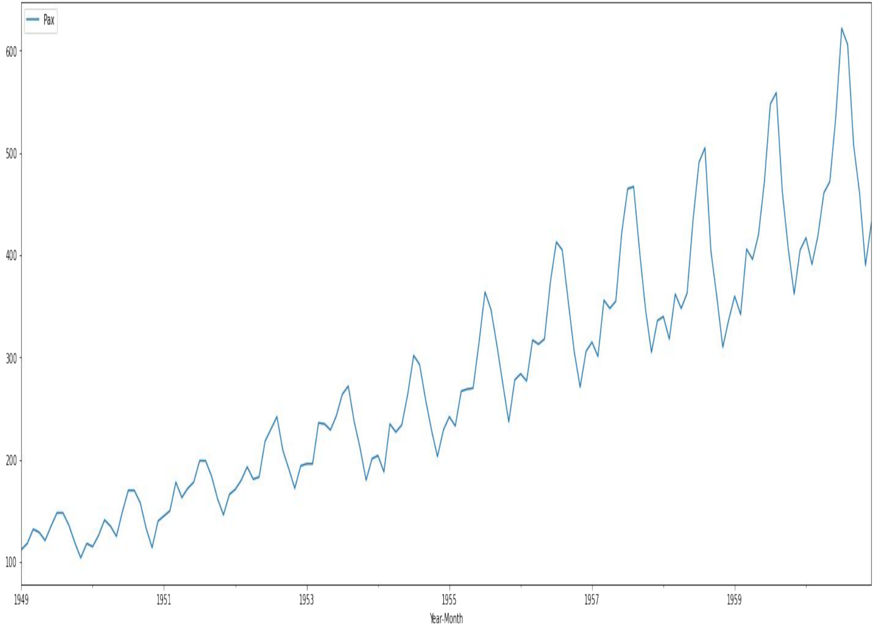


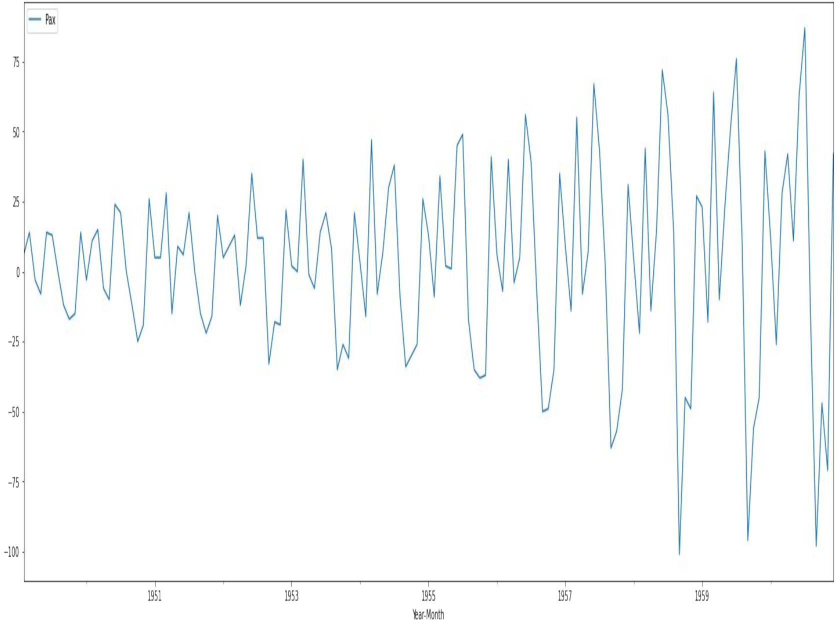
AirTemp TS: seasonal differencing(diff period=12)



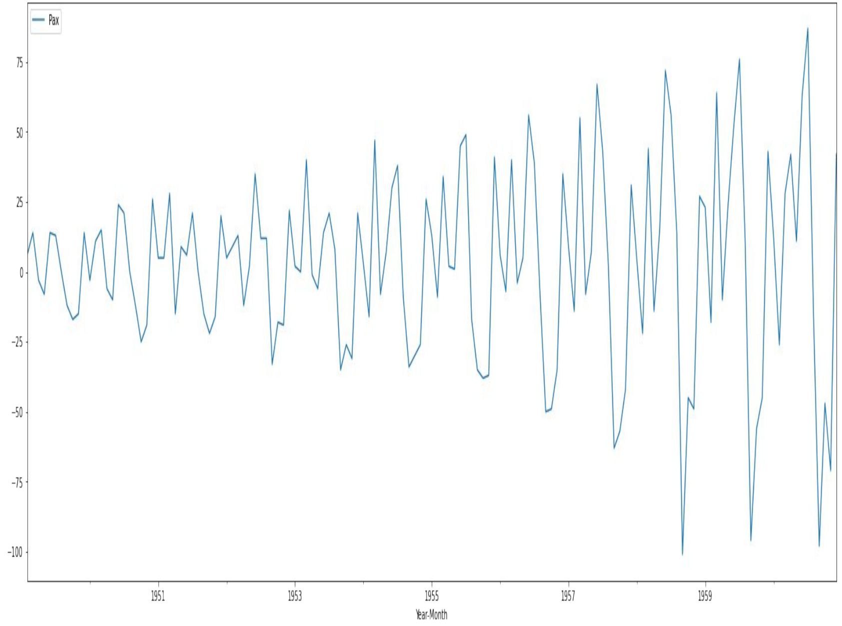


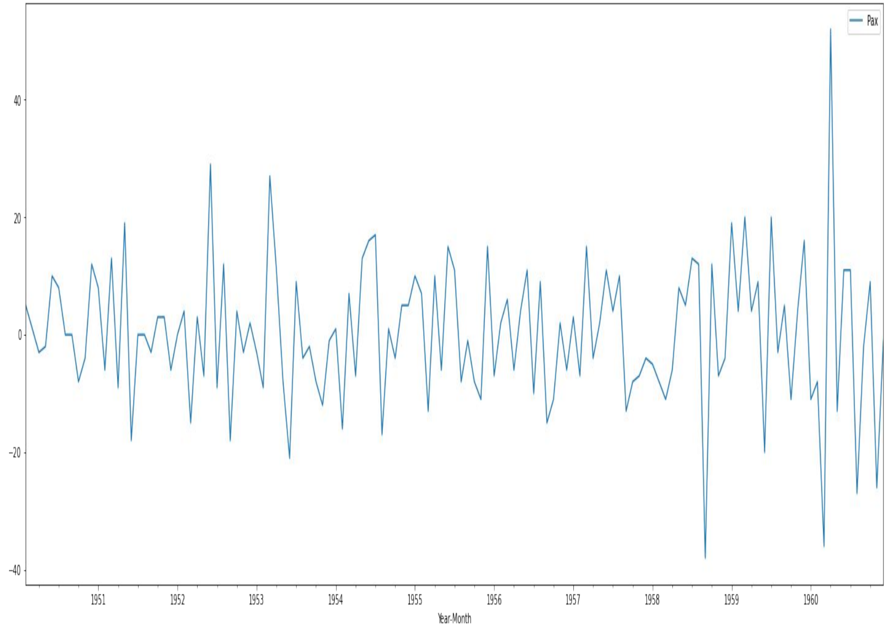
Air Passenger TS: 1st order differencing for trend





Air Passenger TS: seasonal differencing (diff period=12)





Time Series Processes

Autoregressive process(p)

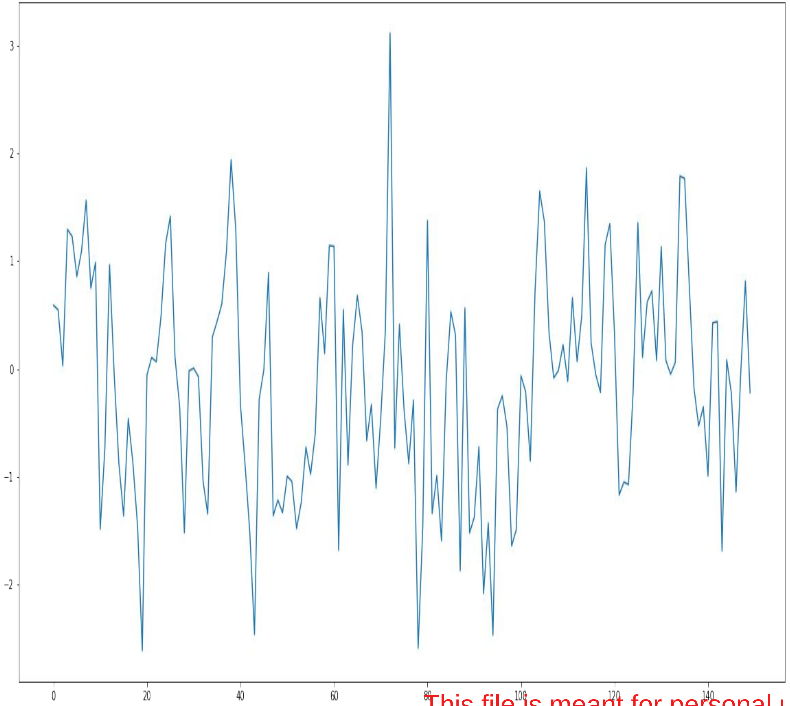
PES UNIVERSITY

- Autoregressive process uses previous time period values to predict the current time period values.
- Autoregressive process is the process denoted by AR(p) where q denotes order of process. The simple autoregressive process of order p can be represented as

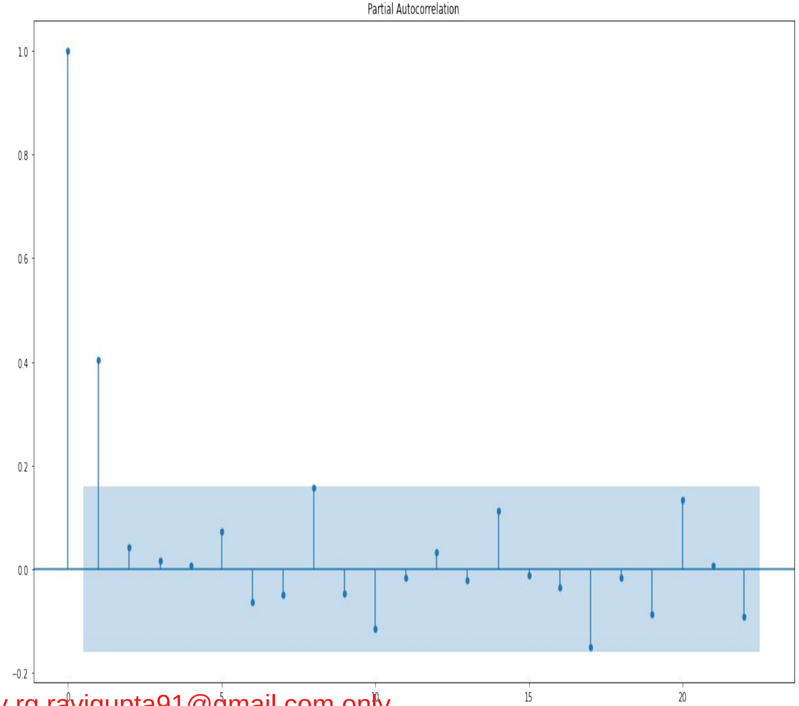
$$y_t = \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \ldots\ldots + \varphi_p y_{t-p} + \varepsilon_t$$

- Where y time series variable, φ are numeric coefficients to be multiply to lagged time series variable.
- ε is the residual term considered as purely random process with mean 0, variance σ^2 and $Cov(\varepsilon_t, \varepsilon_{t-q})$ 0.

AR(1) process

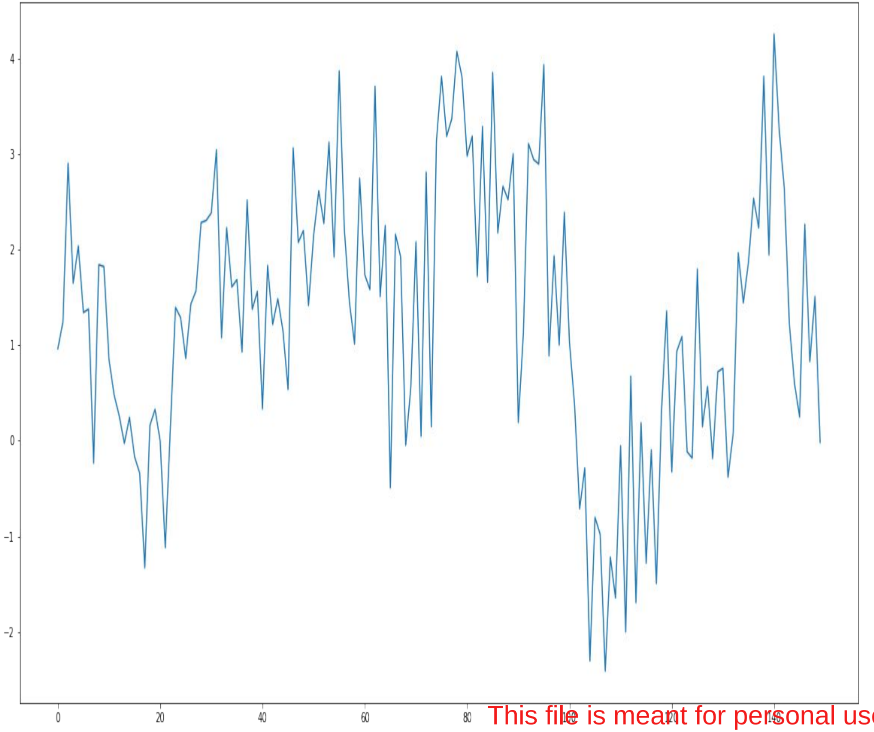


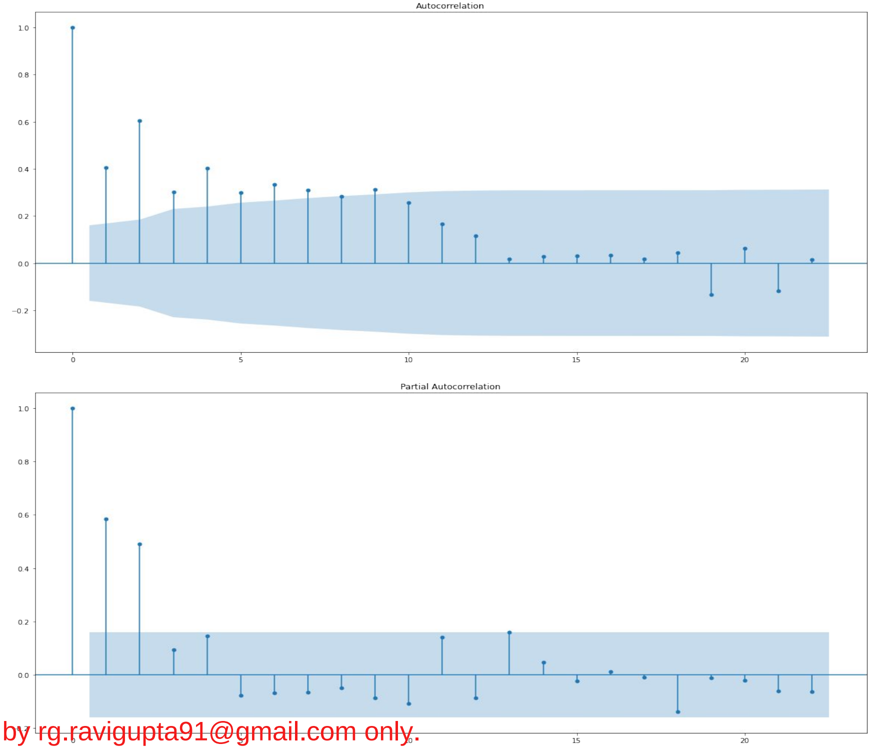
3
2
1
0
−1
−2
0
20
40
60
80
100
120
140



Partial Autocorrelation
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
0
5
10
15
20

AR(2) Process





Moving Average Process(q)

- Moving average process is denoted by MA(q) where p denotes order of process.

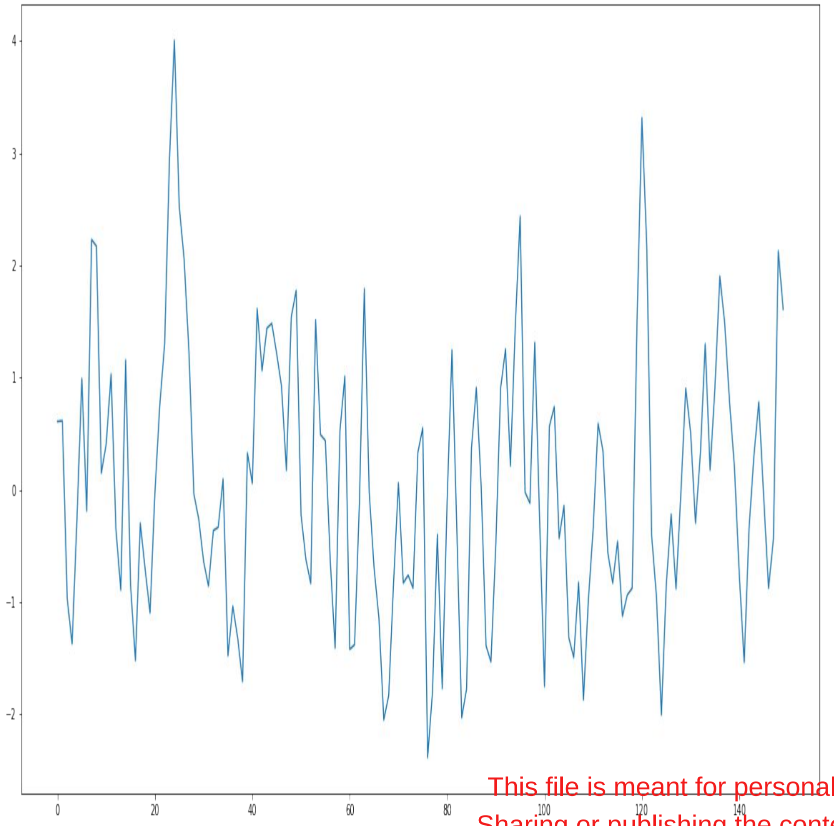
- Moving average process considers past residual values to predict the current time period values.

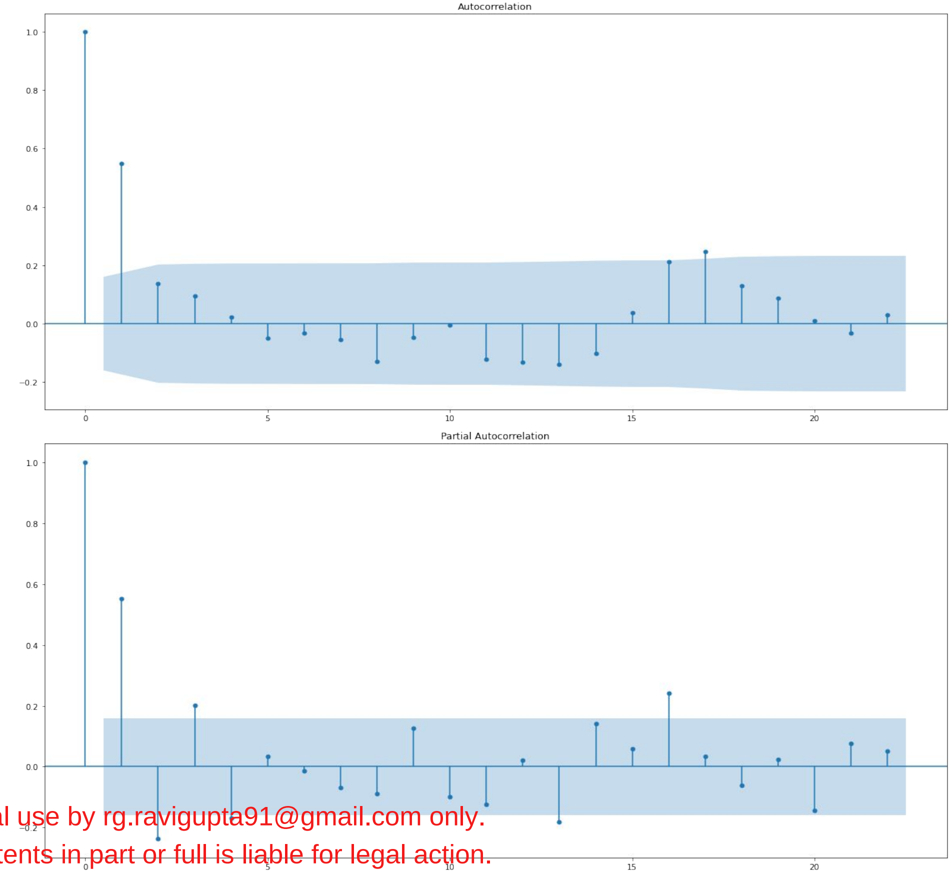
- The moving average process of order q can be represented as

$$y_t = \theta_0 \varepsilon_t + \theta_1 \varepsilon_{t-1} + \ldots\ldots + \theta_q \varepsilon_{t-q} \qquad \text{traditionally } \theta_0 = 1$$

- Where y_t time series variable, θ are numeric coefficients to be multiply to lagged residuals and ε is the residual term considered as purely random

MA(1) process





Autoregressive Moving Average process

- An Autoregressive Moving Average process is a combination of Autoregressive process and Moving average process.

- An Autoregressive Moving Average process estimates the future values considering previous time period values, as well as past errors.

- Autoregressive Moving Average process is denoted ARMA(p, q).

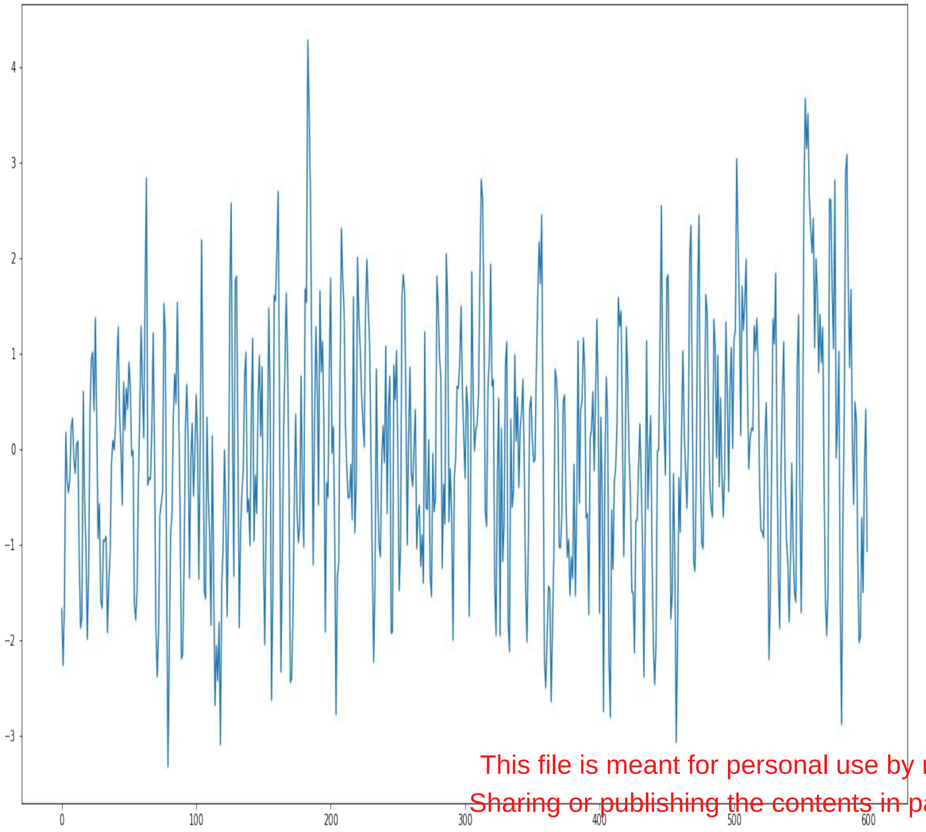
Autoregressive Moving Average process

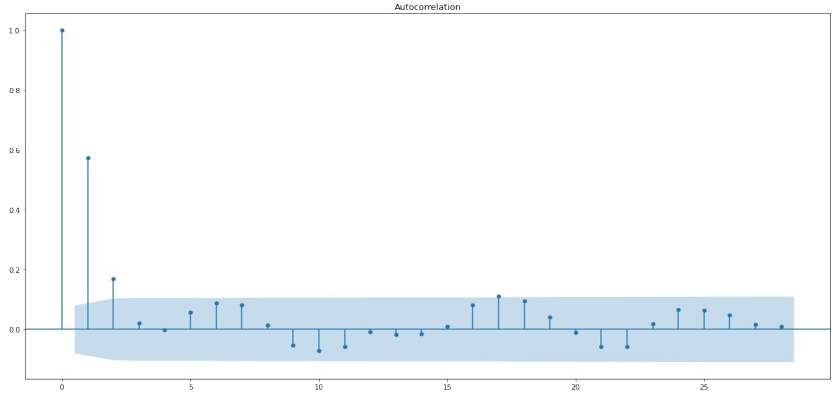
- A simple ARMA process of order (1,1) can be represented as;

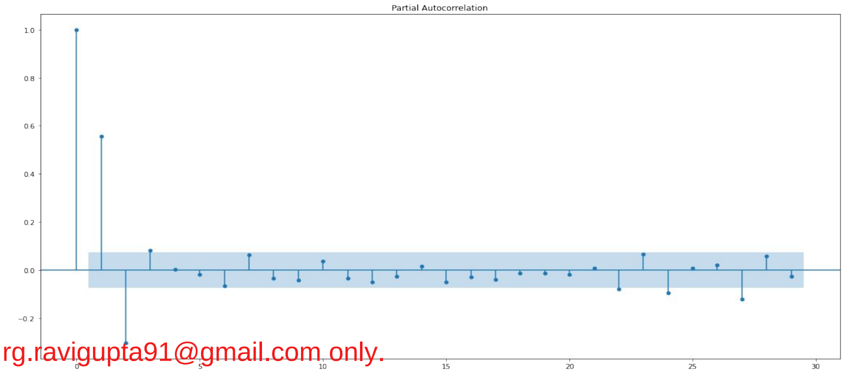
$$y_t = \varphi\, y_{t-1} + \theta\, \epsilon_{t-1} + \epsilon_t$$

- Where y_t and y_{t-1} represents current time period value and previous time period value.
- ϵ_t and ϵ_{t-1} are the error terms for current time period value and previous time period value.
- φ and θ are parameters ranges between range between -1 and 1.
- ARMA process of order (p, q) considers previous values up to p previous periods and residual terms up to q previous periods.

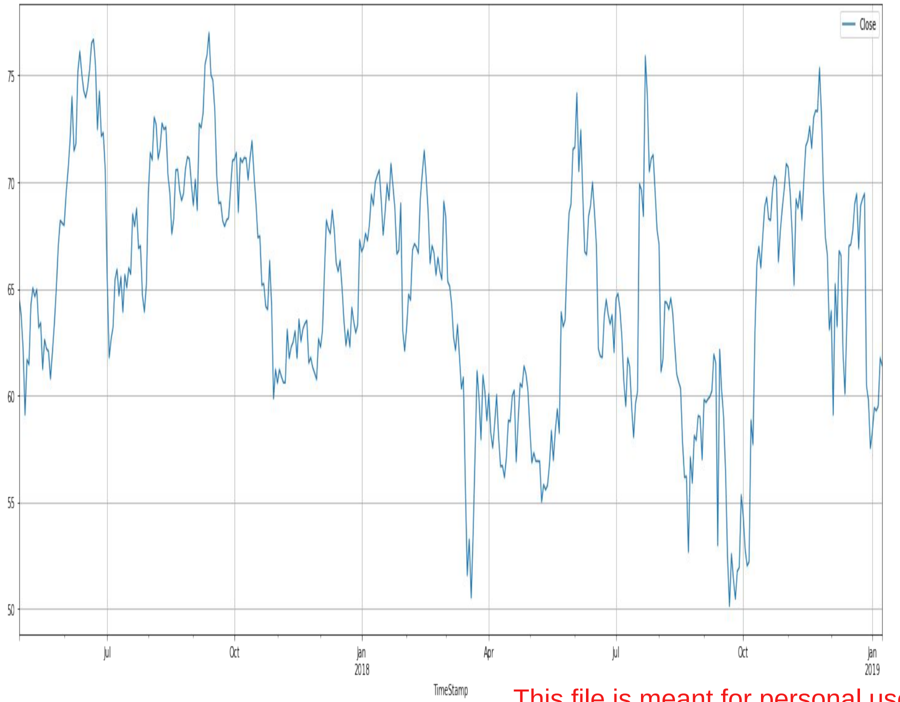
ARMA(2,2) process

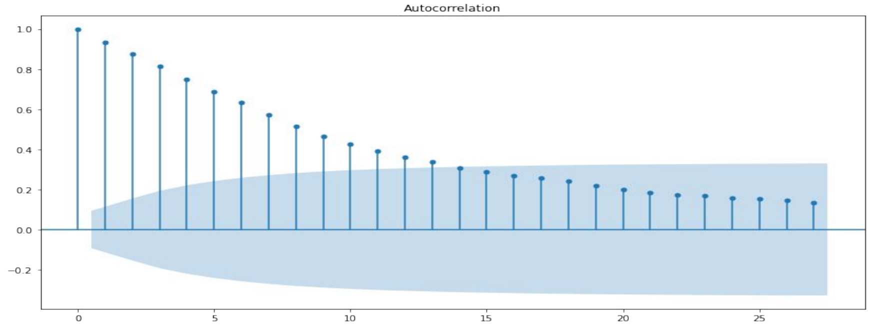


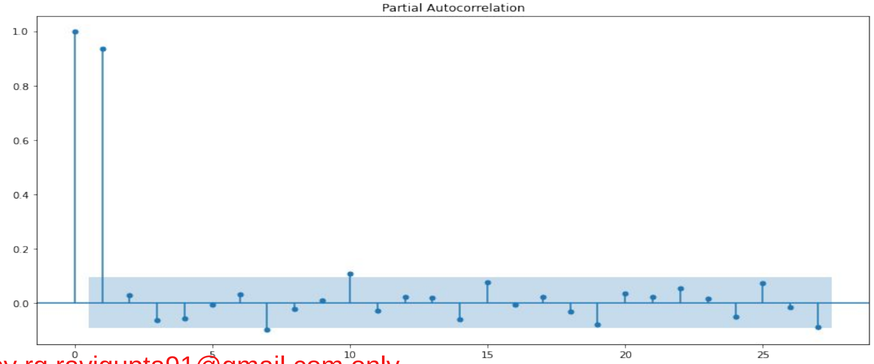




Tesla stock price: ARMA(1,0) process



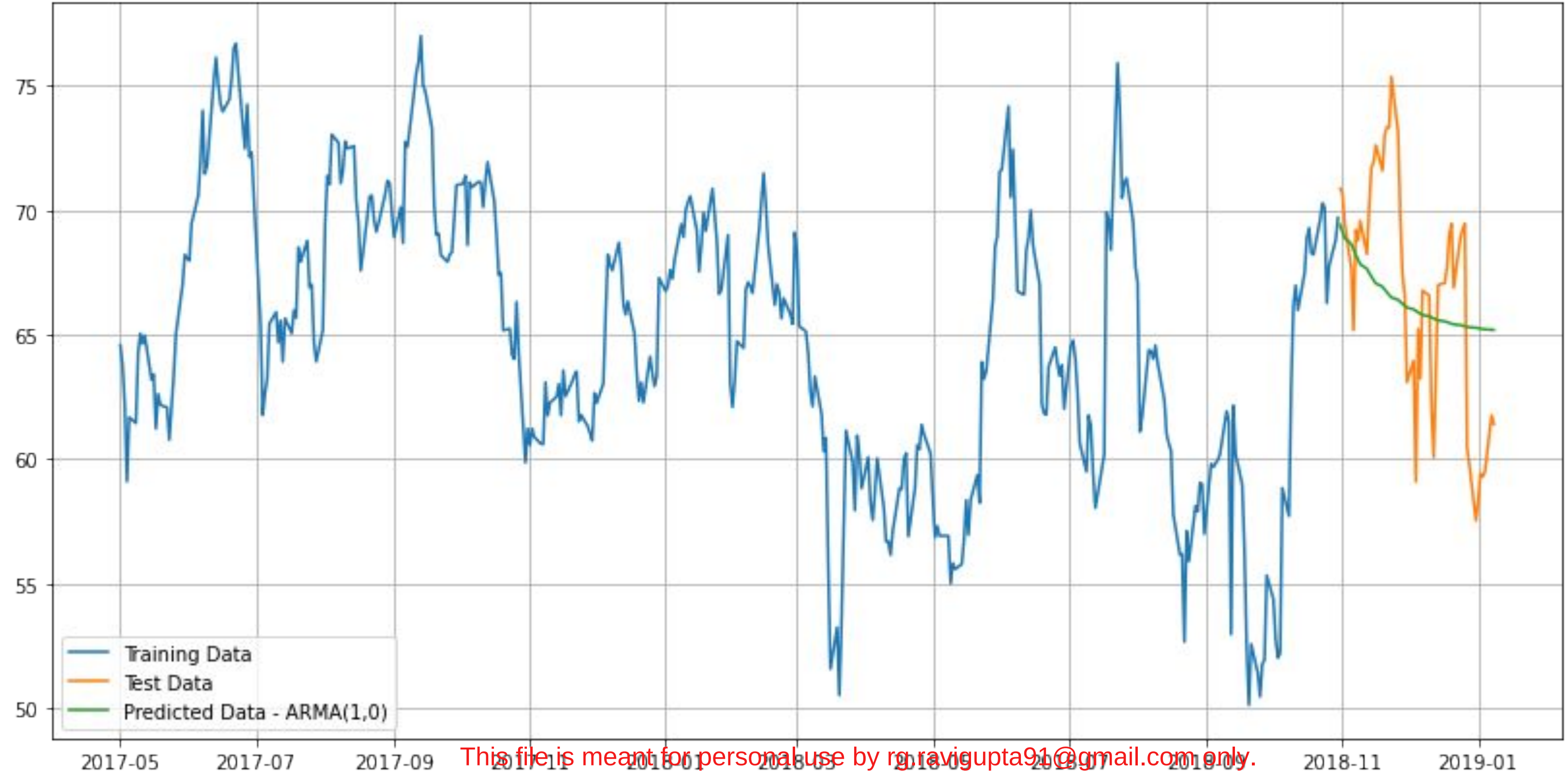




Tesla stock price: AR(1) model summary

```
                              ARMA Model Results
==============================================================================
Dep. Variable:                  Close   No. Observations:                  392
Model:                     ARMA(1, 0)   Log Likelihood                -812.829
Method:                       css-mle   S.D. of innovations              1.919
Date:                Mon, 01 Mar 2021   AIC                           1631.658
Time:                        18:42:19   BIC                           1643.572
Sample:                    05-01-2017   HQIC                          1636.380
                         - 10-30-2018
===============================================================================
                  coef    std err          z      P>|z|      [0.025      0.975]
-------------------------------------------------------------------------------
const          64.9812      1.524     42.646      0.000      61.995      67.968
ar.L1.Close     0.9387      0.017     55.553      0.000       0.906       0.972
                                    Roots
=============================================================================
                  Real          Imaginary           Modulus         Frequency
-----------------------------------------------------------------------------
AR.1            1.0653           +0.0000j            1.0653            0.0000
-----------------------------------------------------------------------------
```

Tesla stock price: forecast using AR(1) model



Forecasting Stationary Time Series

Forecasting Stationary Time Series

- Plot the time series: To observe time series components
- Check for stationarity of time series
- Use ACF/PACF plot to make preliminary choices of process parameters
- Estimate parameters for process
- Fit model to residuals.
- Forecast time series by forecasting residuals.

ARMA(p,q) process Parameter Estimation

Maximum Likelihood Estimator

- Once the order of ARMA model is selected(p,q), we need to estimate the model parameters i.e. ϕ and θ.

- Maximum likelihood estimate is used to best fit the ARMA model for given time series.

- This technique finds the values of the parameters which maximize the probability of obtaining the data that we have observed.

- For ARIMA models, Maximum likelihood estimate is obtain by minimizing,

$$\sum_{t=1}^{n} {\varepsilon_t}^2$$

Summary

- Lag and Differencing
- ACF and PACF
- Stationary process
- Dicky Fuller test
- AR, MA and ARMA process
- Parameter estimators

Thank You