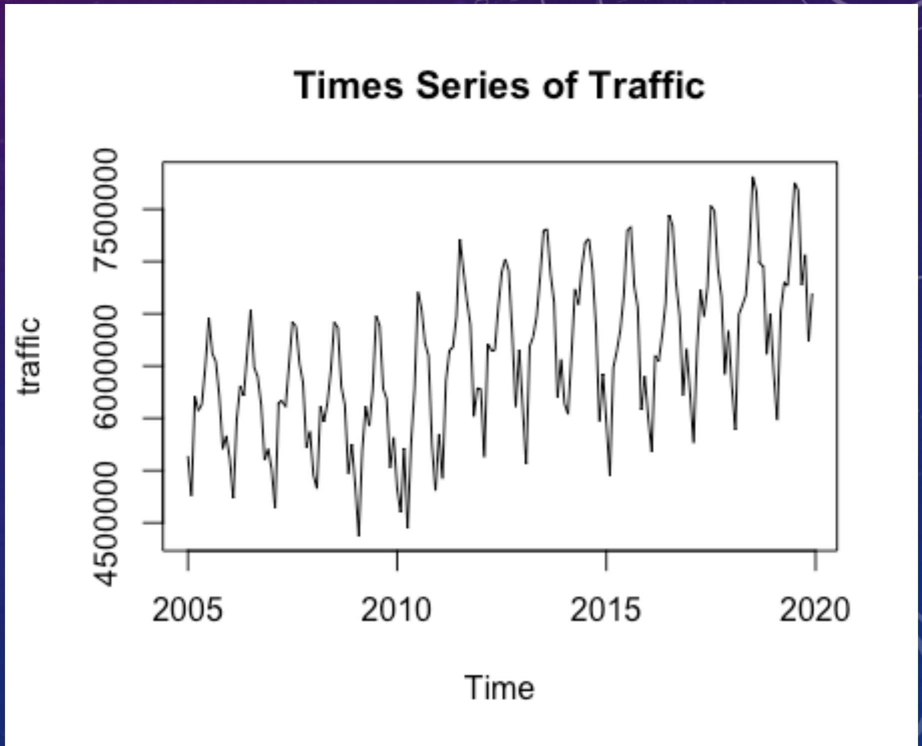
TIME SERIES AND BUSINESS DATA TIME SERIES PROJECT

FORECAST HEATHROW AIRPORT TRAFFIC TIME SERIES

DESCRIPTION OF THE TIME SERIES

- The time series includes the data of monthly traffic of Heathrow Airport, London, The United Kingdom
- From period of January 2005 to December 2019, in total 180 observations
- The data table composes of 3 columns: year, month and traffic of the airport
- The graph of time series shows that it has an increasing trend, changing mean and variance against time

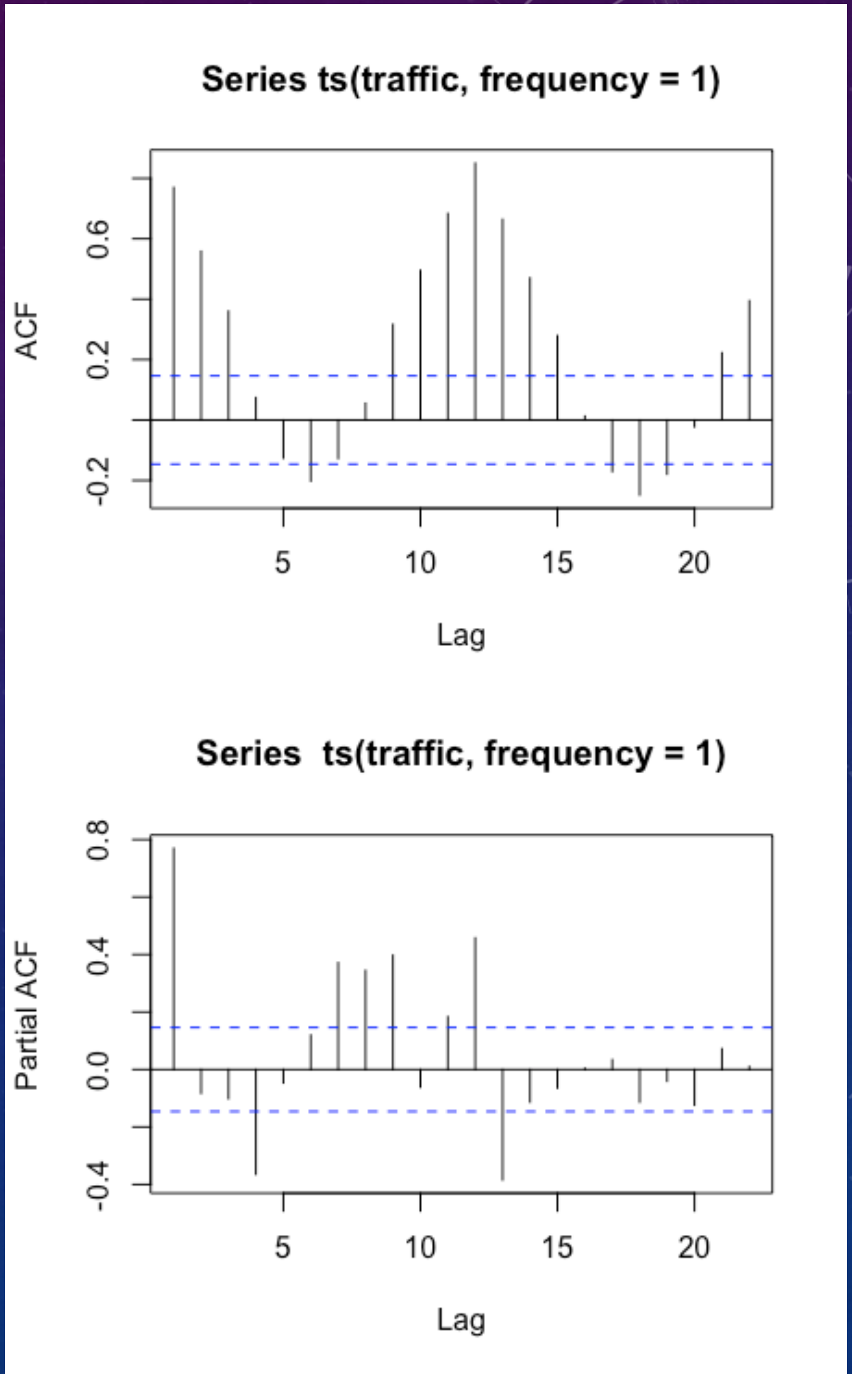
➢ Hence the time series is not stationary



STEP 0: STATIONARIZATION OF DATA

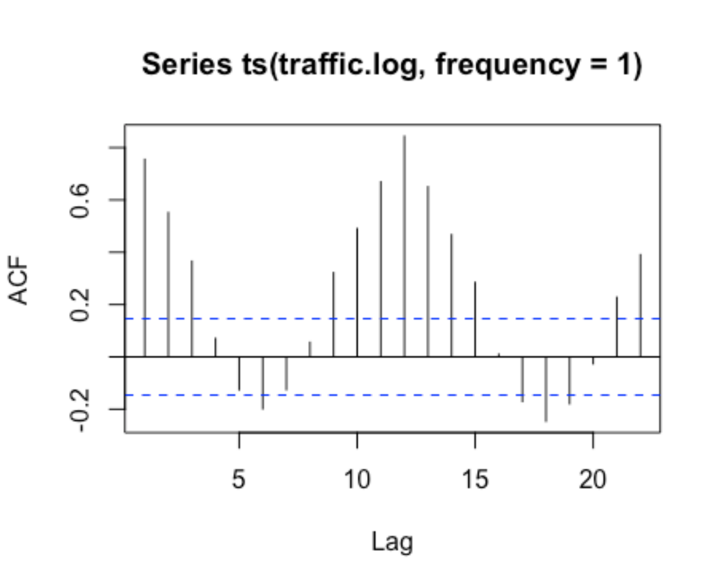
- The upper graph, autocorrelation function (ACF) of original data, shows that the series will not decay to zero
- The lower graph, partial autocorrelation function (PACF) of original data, also shows similar results

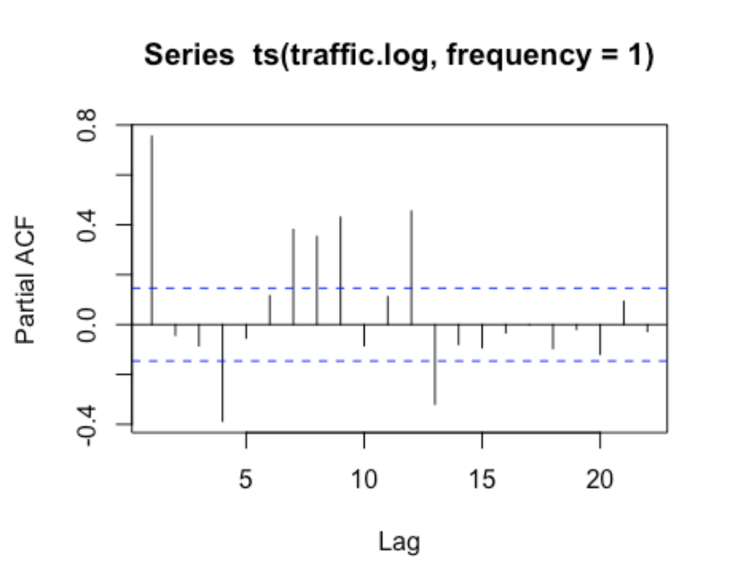
➢ It is necessary to perform a log transformation of the time series to try to reach stationarity



LOG TRANSFORMATION

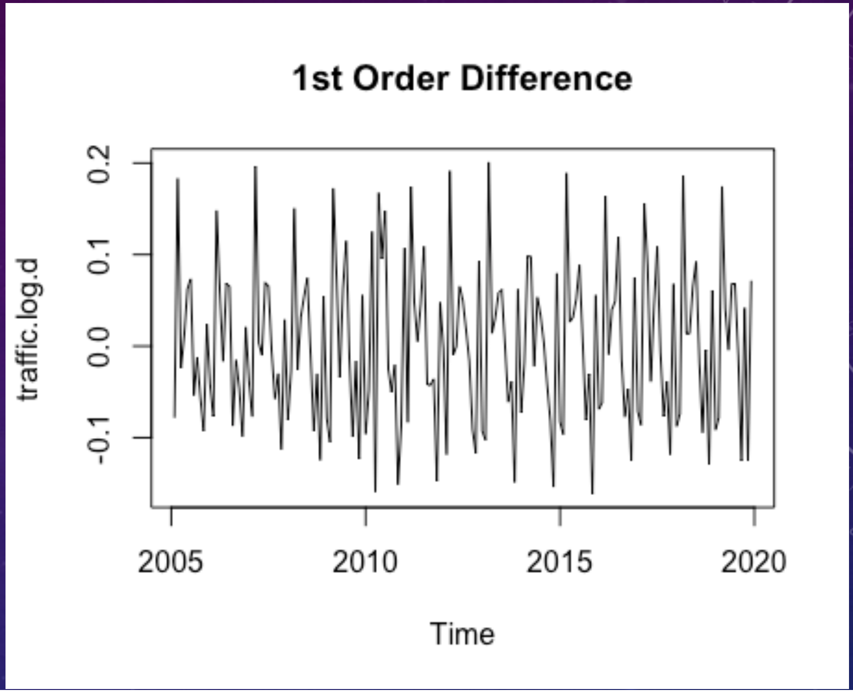
- The graph of log transformation (upper one) shows that the trend still remains
- The ACF and PACF graphs (below ones) yield similar results
- Need to do further transformations

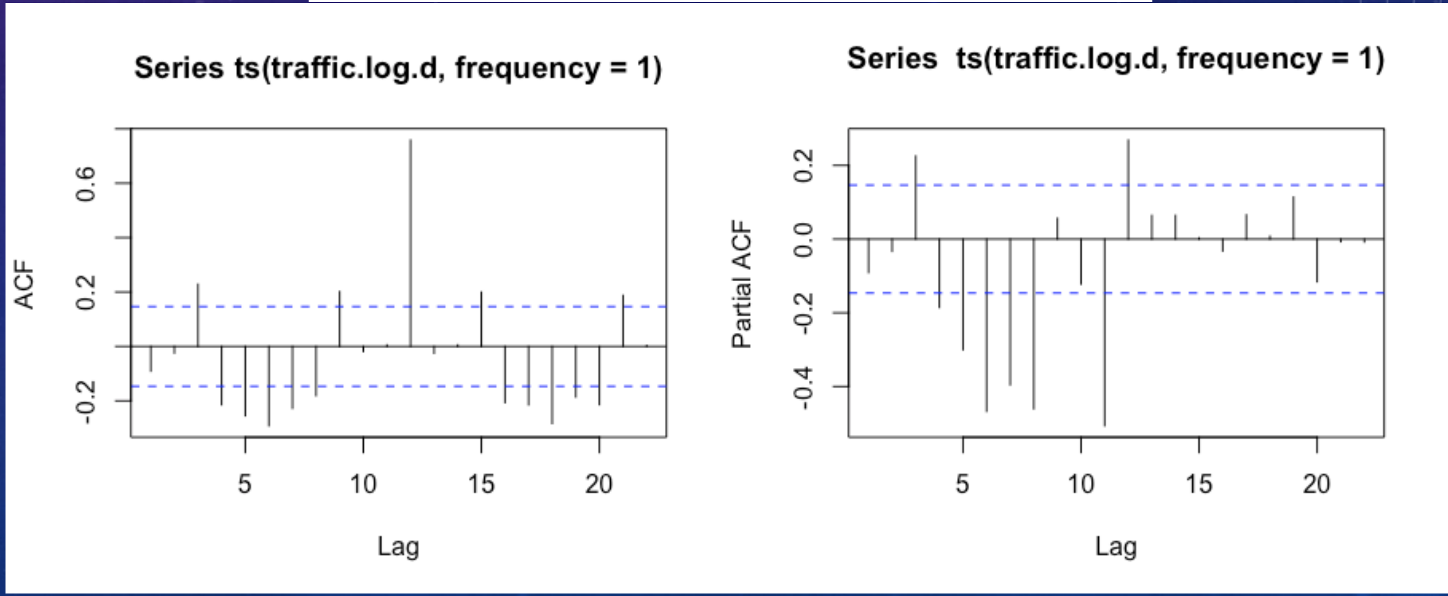




1^{ST} ORDER DIFFERENCE

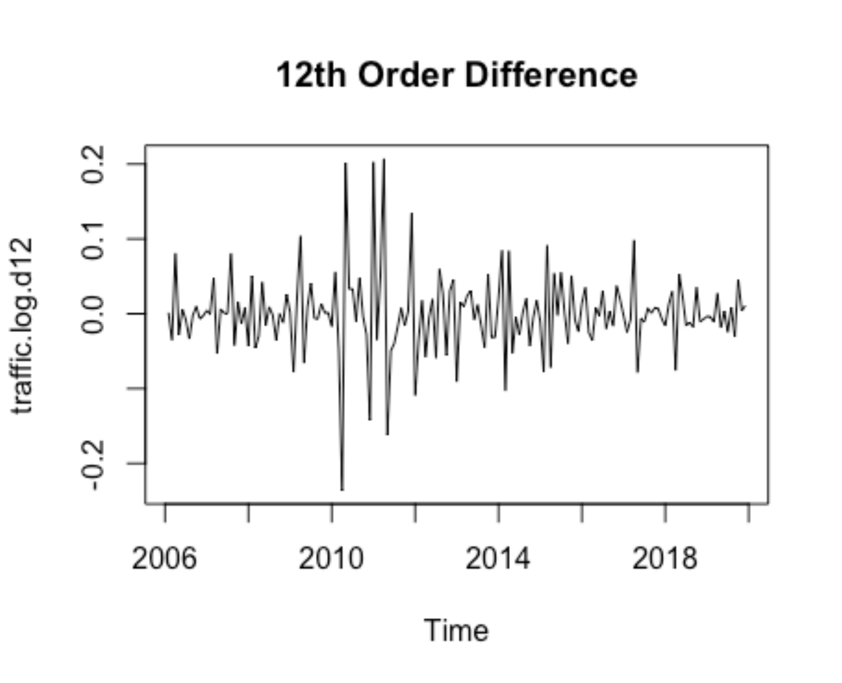
- 1^{st} order difference is to remove the trend
- The graph of 1^{st} order difference (upper one) indicates that the series is more stationary
- Yet we can clearly see the series is seasonal
- The ACF and PACF graphs (below ones) are more inclined to decay than before transformation.

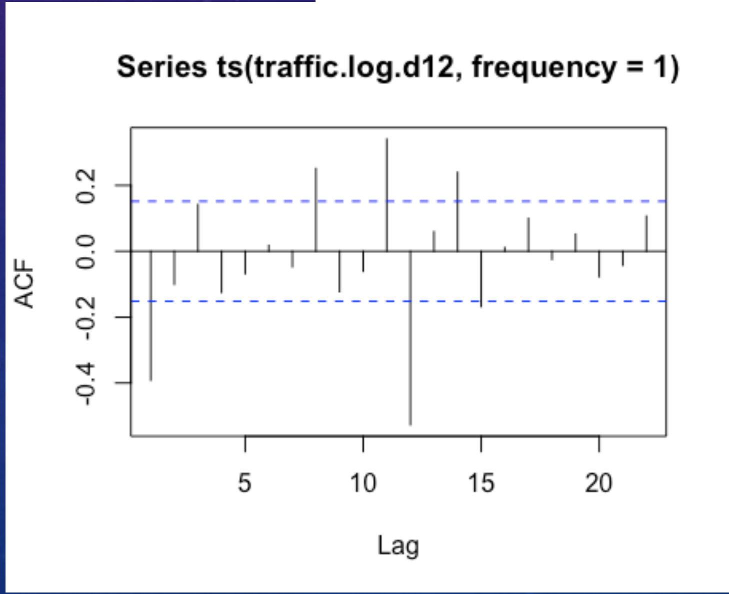


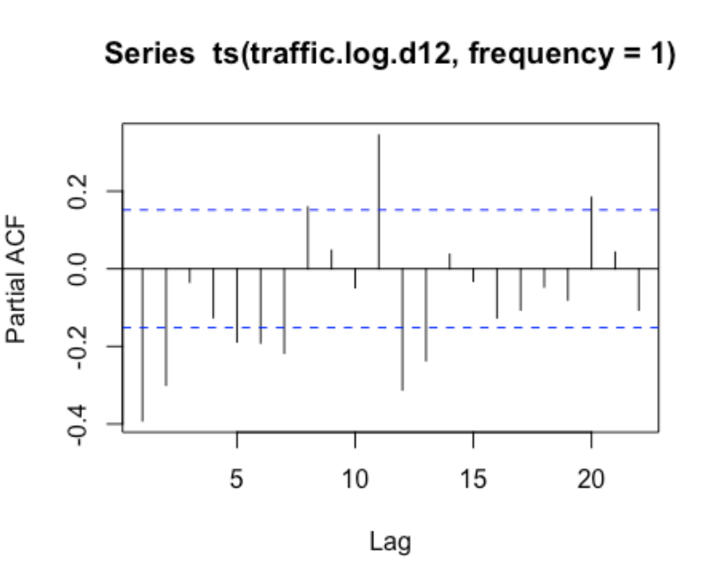


12^{TH} ORDER DIFFERENCE

- 12^{th} order difference is to remove the seasonality
- The graph of 12^{th} order difference (upper one) has no seasonality anymore
- The ACF and PACF graphs (below ones) decay to zero faster than before this transformation

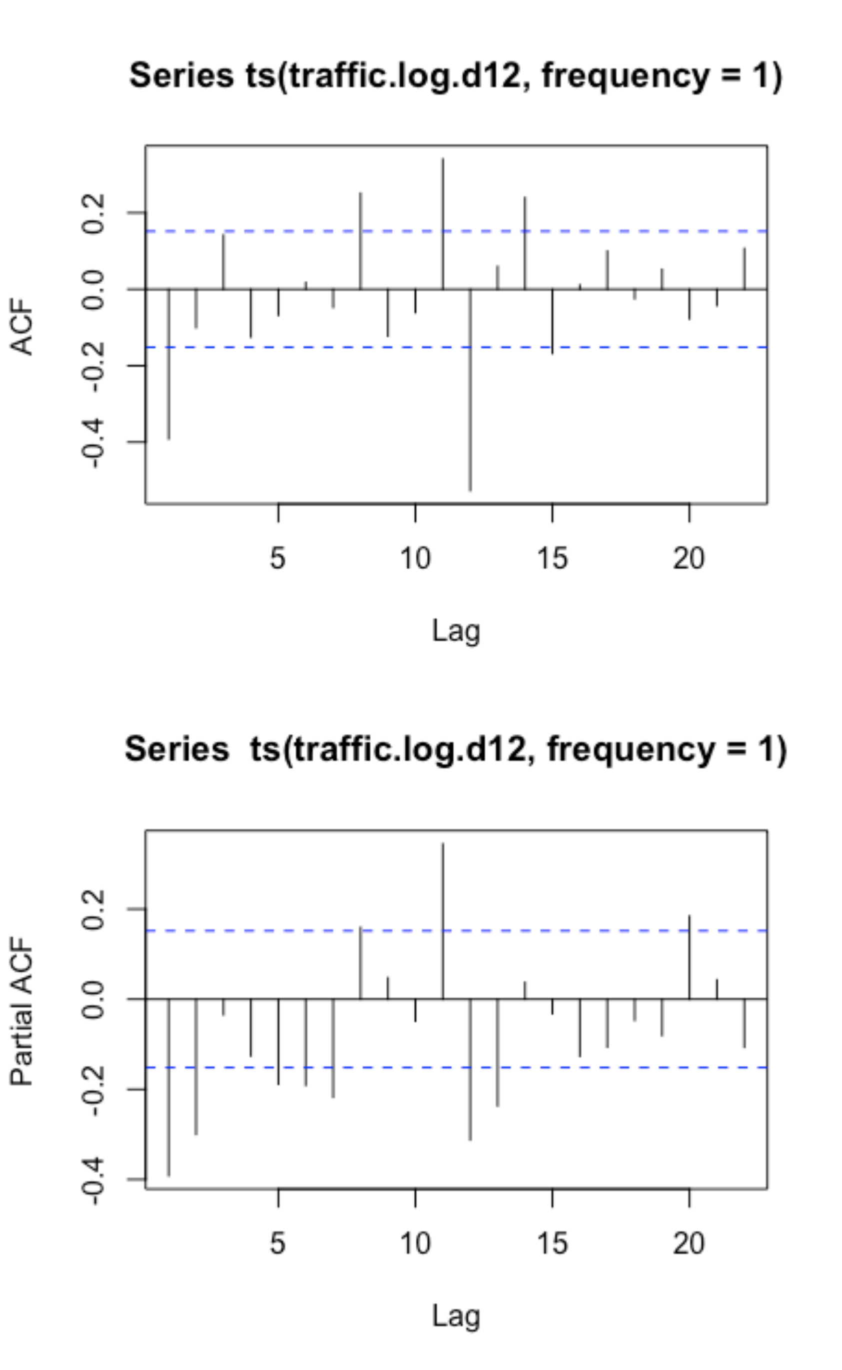






STEP 1: IDENTIFICATION OF ORDERS

- From the graph of ACF (above), we can clearly see that q = Q = 1
- From the graph of PACF (below), we can clearly see that p = P = 2
- From both graphs, we can deduce s = 12

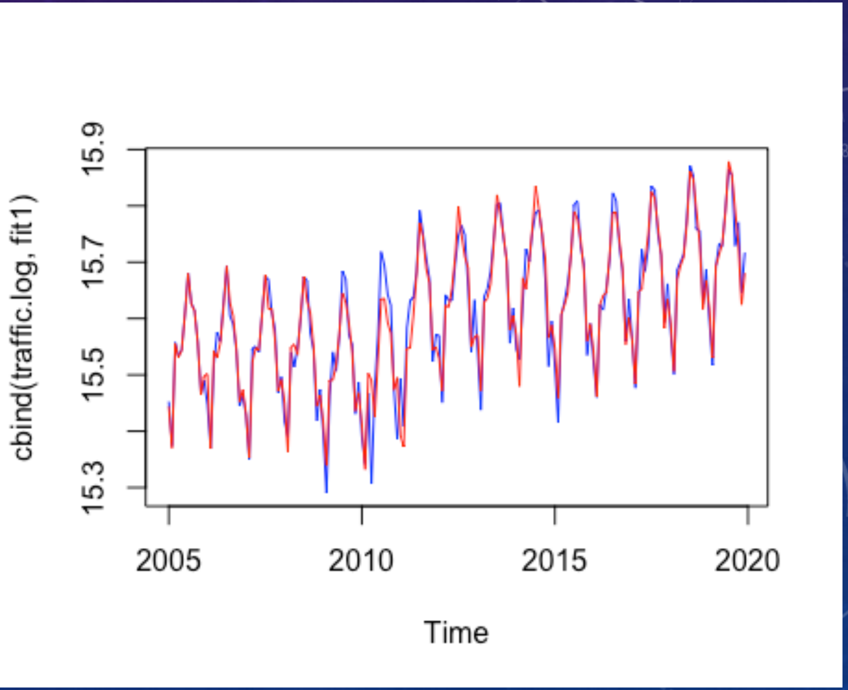


STEP 2: ESTIMATION OF ARMA COEFFICIENTS

- From the orders of the model, we can generate the coefficients of model 1:

ar1	ar2	ma1	sar1	sar2	sma1
0.1929	0.0325	-0.7679	0.0381	0.0594	-0.9999

- A graph of fitted values (in red) is also drawn to predict values and compare with the model



STEP 3: MODEL CHECKING

- Calculate the p values of the coefficients to see whether they are significant

ar1	ar2	ma1	sar1	sar2	sma1
0.068283341	0.714612055	0.000000000	0.636177617	0.449294116	0.003862191

- ar2, sar1 and sar2 are insignificant
- Remove them and re-estimate the model again to generate model 2
- p = 1, P = 0, q = Q = 1, d = D = 1, s = 12
- The coefficients would be:

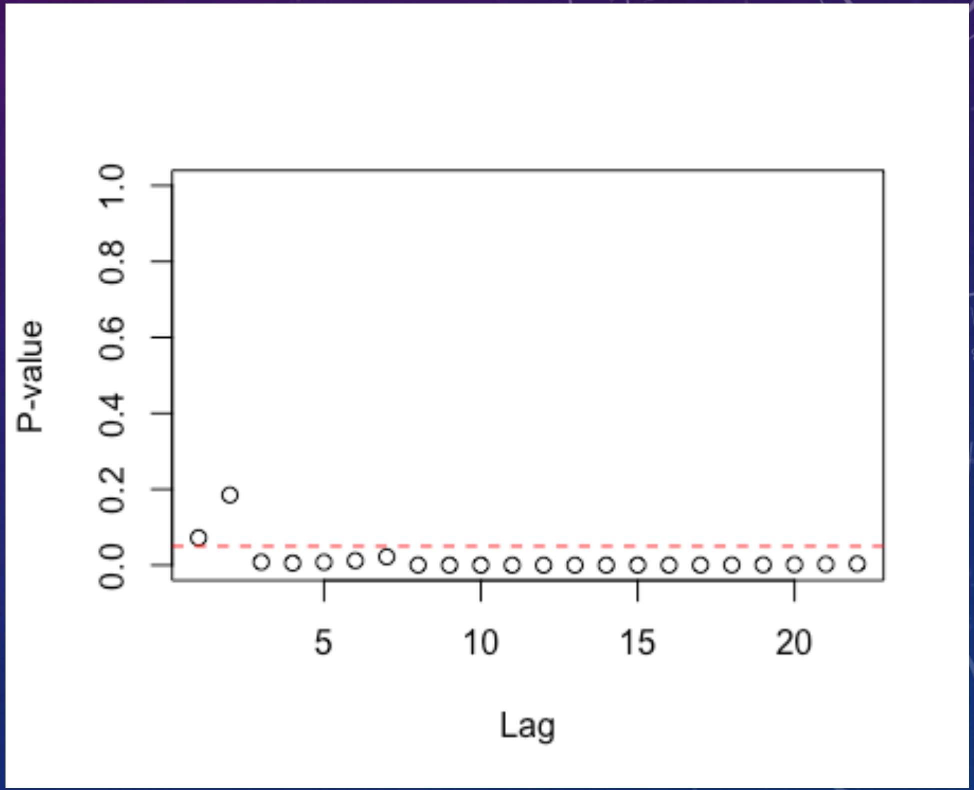
ar1	ma1	sma1
0.1821	-0.7500	-0.9121

STEP 3: MODEL CHECKING

- Check the p values again to see whether they are significant

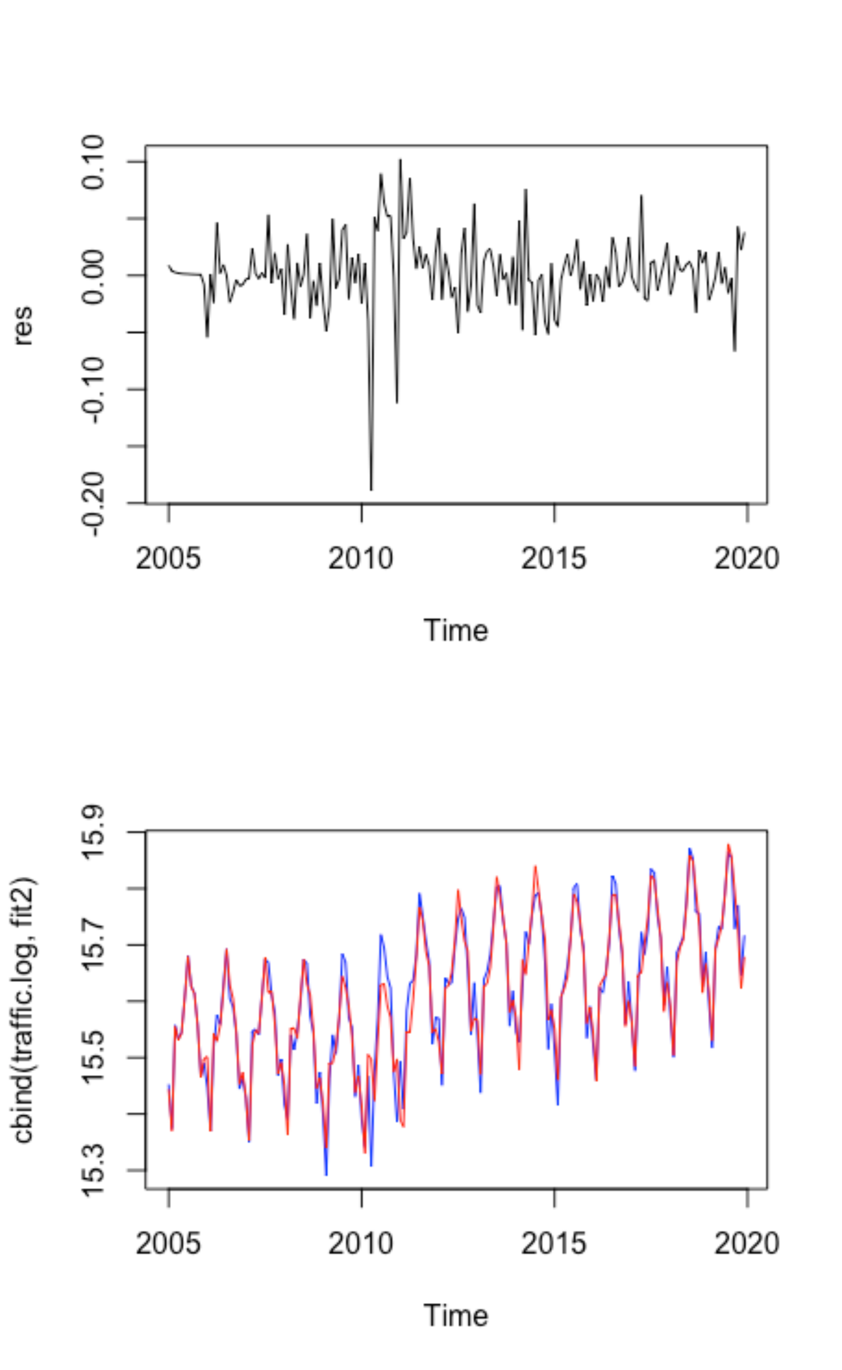
ar1	ma1	sma1
1.090937e-01	0.000000e+00	5.107026e-15

- ➢ They are significant
- Perform McLeod Li test to check the model's homoscedasticity
- From the graph, most of the p values are under the red line of 5%
- ➢ Reject the null hypothesis, thus the model has heteroscedasticity
- ➢ Maybe due to outliers of data



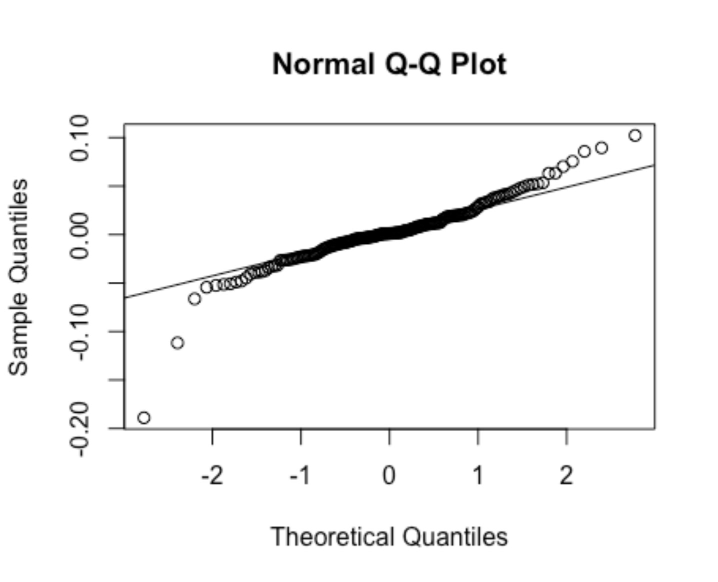
STEP 3: MODEL CHECKING

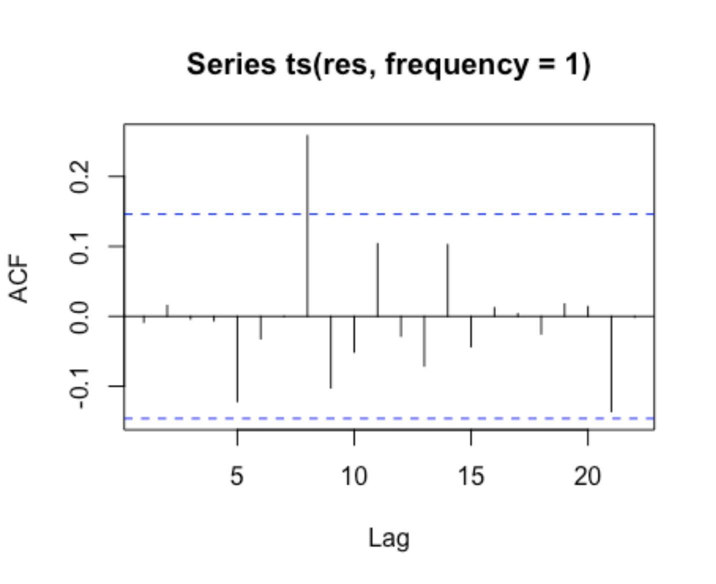
- Plot the graph of residuals (upper graph)
- ➢ We can clearly see that there are outliers around 2010 and 2011
- Plot its fitted values (lower graph)
- The line in red is the fitted values
- ➢ It is obvious that the period where outliers are located have caused bigger difference between residuals and fitted values
- Jarque Bera Test obtains p-value < 2.2e-16
- ➢ The residuals do not have normal distribution
- ➢ Maybe also due to the outliers

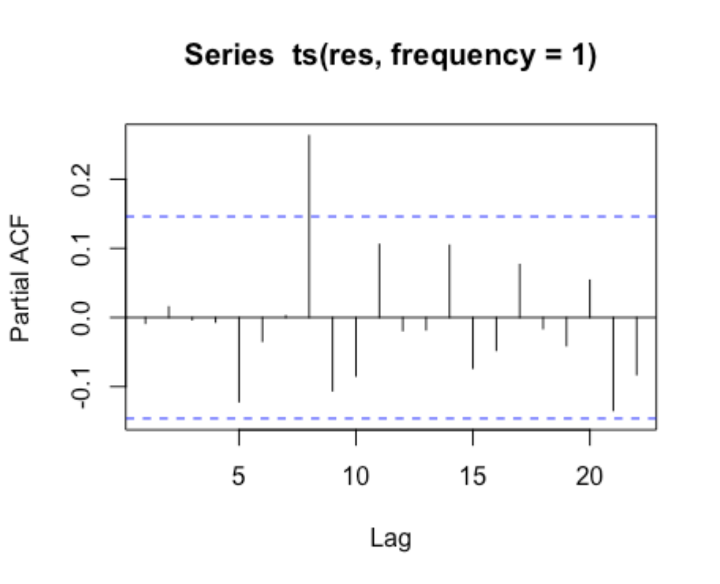


STEP 3: MODEL CHECKING

- From the graph of QQ plot (upper one), we can conclude that the plot is not linear
  - ➢ The residuals are not normally distributed as there are obvious outliers
- The lower graphs of ACF and PACF of residuals show that there is no autocorrelation as there is no significant coefficient
- The result from Box Ljung test further strengthens the above result:
  - ➢ p-value = 0.2335







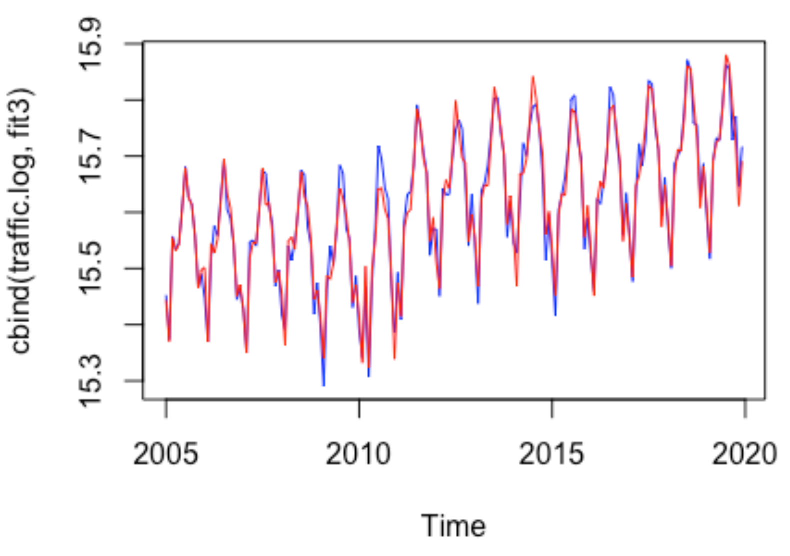
STEP 3: MODEL CHECKING

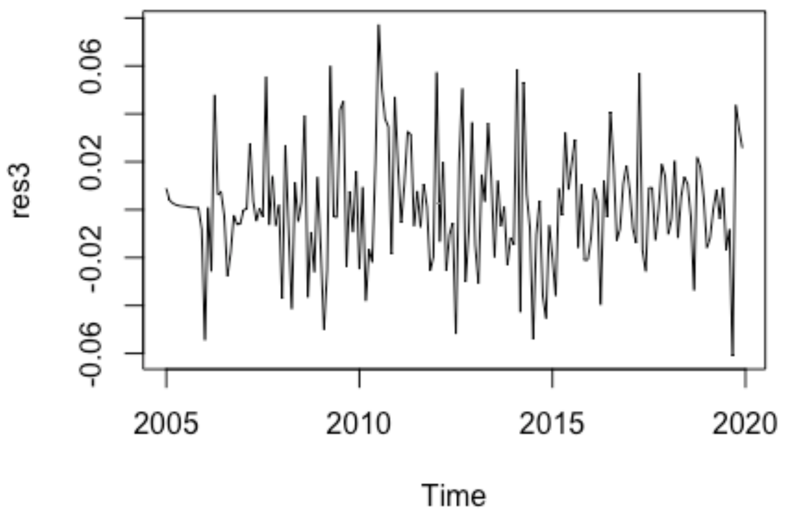
- The details of outliers:

Date	Value	Traffic
2010.250	-5.650490	4446530
2010.917	-3.339927	4809195

- To deal with the outliers, we assign dummy variables to replace outliers and create model 3 for estimation
- Again we generate the coefficients below, fitted values and residual analysis respectively

ar1	ma1	sma1	xreg
8.310479e-01	4.703146e-06	0.000000e+00	0.000000e+00





STEP 3: MODEL CHECKING

- Shapiro Test result: p-value = 0.02926
- Jarque Bera Test result: p-value = 0.1637
- Without outliers, the QQ plot of residuals shows that it is linear and normally distributed

➢ We can accept that the residuals are normally distributed

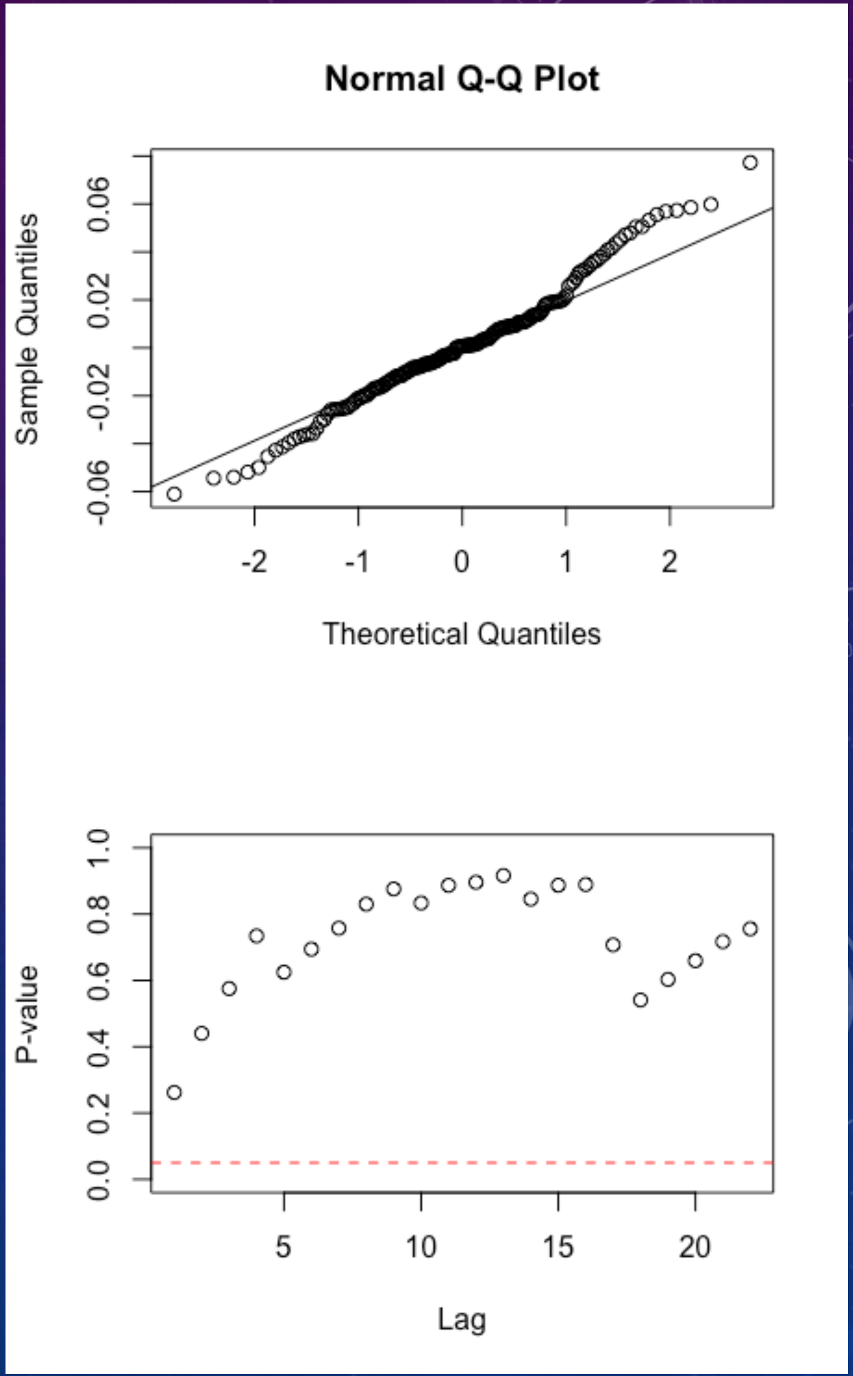
- All p values of the graph of McLeod Li test are above 5%

➢ Hence the residuals have homoscedasticity

- The AIC and SBC of all models show that model 2 is the best fit

➢ It has the lowest values of both AIC and SBC

Model	AIC	SBC
Model 1	-626.2883	-604.4624
Model 2	-631.7395	-619.2676
Model 3 (with dummy values)	-731.0842	-715.4942



STEP 4: PREDICTION

- In-sample and out-of-sample analysis:

In-sample data	Out-of-sample data
Jan 2005 – Jan 2015	Feb 2015 – Dec 2019

- Accuracy:

	ME	RMSE	MAE	MPE	MAPE	ACF1
Test set	0.01729778	0.03286758	0.02643252	0.1096912	0.1682953	0.376278

➢ ME, RMSE, MAE are relatively small, showing that the model is at least 96.8% accurate which is a good model

➢ MPE and MAPE are higher but still the majority of the model is accurate

➢ ACF at lag 1 has the correlation between first and next point is around 0.37, showing that there is an exponential decrease

STEP 4: PREDICTION

- Plot the graph of quality of fit for examination
- Calculate the proportion of points in the confidence bound to see how good the fit is: 86.11%

➢ The fit is good

- The predicted values for 3 months are shown in red
- The dotted lines show the upper and lower confidence level