1. FullDate: Date in format yyyy-mm-dd  hh:mm:ss
2. ElecPrice: Prices of electricity charged in Kw per hour
3. Gas Price: power generated was done via a gas power plant, gas price is the price of gas at time-T
4. SysLoad: Total Load on system or Power consumed at time-T
5. Tmax: maximum temperature at time-T

**Various Forecast Performance Measures**

As we have emphasised in previous modules, it is essential to correctly evaluate our models with the appropriate metrics. In time series forecasting, there are a number of measures that can be used to evaluate how well a model performs. We discuss some of these below.

**Mean Absolute Error (MAE):**

This is a measure of the forecast error obtained by taking the average of the absolute errors of the forecasted values where the error is the difference between the forecasted value and the true value. Simply put, it tells the average of the error that should be expected from a forecast. The MAE is robust to**o**utliers and scale-dependent error so it cannot be used to compare different series.

\textmd{If}~ {y\_i} ~\textmd{and}~\hat {y\_i} ~\textmd{represent the actual value and forecasted value} \newline \textmd{respectively,} \textmd{then} \newline \huge MAE~=~{1\over n}\sum^{n}\_{i=1} |y\_i~-~\hat y\_i|If *yi*​ and *yi*​^​ represent the actual value and forecasted valuerespectively,then*MAE* = *n*1​∑*i*=1*n*​∣*yi*​ − *y*^​*i*​∣

**Mean Absolute Percentage Error (MAPE):**

It is a well known measure used to determine the accuracy of a forecasting method. The accuracy is calculated as a percentage such that the average of the difference between the actual value and the forecast value divided by the actual value. The average of the sum of the absolute value of this calculation for every data point forecasted is multiplied by 100 to obtain a percentage error. MAPE can be easily interpreted, scale independent and should not be used with low-volume data. In addition, it should not be used when the values are zero or close to zero because it results in undefined and indefinite values.

\huge MAPE ~=~({1\over n} \sum^{n}\_{i=1} |{y\_i~~ -~~\hat y\_i\over y\_i}| ~\* ~100)*MAPE* = (*n*1​∑*i*=1*n*​∣*yi*​*yi*​  −  *y*^​*i*​​∣ ∗ 100)

**Root Mean Square Error (RMSE):**

The RMSE calculates square root of the average of the squared errors  (i.e. the square of the difference between the actual and forecasted value).  This measure tends to exaggerate large errors by giving a relatively high weight to large errors because it takes the square root of the average squared errors. The Mean Squared Error (MSE) is simply the average of the square of the errors.

\huge RMSE~=~\sqrt{{1\over n}\sum^{n}\_{i=1}(y\_i~-~\hat y\_i)^2}*RMSE* = *n*1​∑*i*=1*n*​(*yi*​ − *y*^​*i*​)2

**Time Series Forecasting Using Stochastic Models**

There are several models that can be used to forecast time series, some which we will discuss in this section.

1. Autoregressive Moving Average (ARMA) Models
2. Autoregressive Integrated Moving Average (ARIMA) Models
3. Seasonal Autoregressive Integrated Moving Average (SARIMA) Models

**Autoregressive Model (AR):**

This is a linear combination of previous values of the variable being forecasted. It can be likened to a linear regression of the observations in the current series against the past values from that same series. It is referred to as autoregressive because it is a regression of the variable against itself. The number of past time steps to be inputted is determined by a single argument, p.

**Moving Average Model (MA):**

MA models consider past residuals. Instead of using past values as in AR models, MA models use previous forecast errors to model the next step in the sequence. The order of the model is specified with the q parameter as in MA(q). While this might be mixed up with the moving average described above, the MA model is used to forecast future values while the smoothing technique allows for estimating trend cycles of previous observations.

**Autoregressive Moving Average (ARMA):**

As suggested from the name, ARMA(p, q) is a combination of the AR and MA models. It models the future values in the series as a linear combination of the observations and the errors of previous time steps.

**Autoregressive Integrated Moving Average (ARIMA):**

Although very similar to ARMA, ARIMA includes a differencing phase called integration where the series is made stationary.  ARIMA(p,d,q) includes a third parameter d which represents the number of non-seasonal differences required to make the data stationary. A non-seasonal time series that is not a random white noise and shows patterns can be modeled with an ARIMA model.

**Seasonal Autoregressive Integrated Moving Average (SARIMA):**

This is an extension of the ARIMA model performed at the seasonal level. It is represented as ARIMA(p,d,q)(P, D, Q)m where the new variables (P, D, Q) represent p,d,q for the seasonal part of the series and m represents the number of periods in each session. SARIMA factors in seasons by differencing the current value and its value in the previous season.

**Stationarity Analysis, Autocorrelation and Partial Autocorrelation Functions (ACF & PACF)**

Having discussed the importance of using a stationary time series in forecasting, we now identify methods that can be used to test for stationarity. Visualisations, statistical tests and summary statistics are commonly used to check that a series is stationary.

* Visualisations: As done when exploring data for data analysis tasks, visualising datasets tend to give a quick view to patterns present. Although it is not a dependable method, by plotting time series data, it is possible to visually determine if there is a known stationarity property present. Plots can show an obvious lack of trend and seasonality in the series.
* Statistical tests: These can be used to check if the expectations of a stationary time series are met. An example is the unit root test used to check for the presence of a trend in a series if it has a unit root. For the test, the null hypothesis H0, proposes that the series is non-stationary (time-dependent) because it has a unit root. If the null hypothesis is rejected, the alternate hypothesis H1 proposes that the series is independent of time and does not have a unit root hence, it is stationary. As in statistical tests, a threshold is set for the p-value to reject the hypothesis. If a threshold is set at 5%, then a p-value > 0.05 will fail to reject the null hypothesis suggesting that the time series is non-stationary however, with a p-value <= 0.05, we can reject the null hypothesis and conclude that the series has no unit hence, it is stationary. <br><br>The Augmented Dickey-Fuller (ADF) test is a well known unit root statistical test that allows for higher order autoregressive processes. It is an extension of the Dickey-Fuller (DF) test used for more complex and larger time series models. In python, the adfuller() function used for the test can be found in the statsmodel package. Another unit root test is the KPSS test. With this test, the null and alternate hypothesis are opposite of the ADF test. Simply put, for the KPSS test, the null hypothesis H0 suggests that the series is stationary while the alternate hypothesis suggests that the series is non-stationary.
* Summary statistics: By splitting the series into different groups and comparing the mean and variance across, if there is a significant difference, there is a high chance that the series is non-stationary.

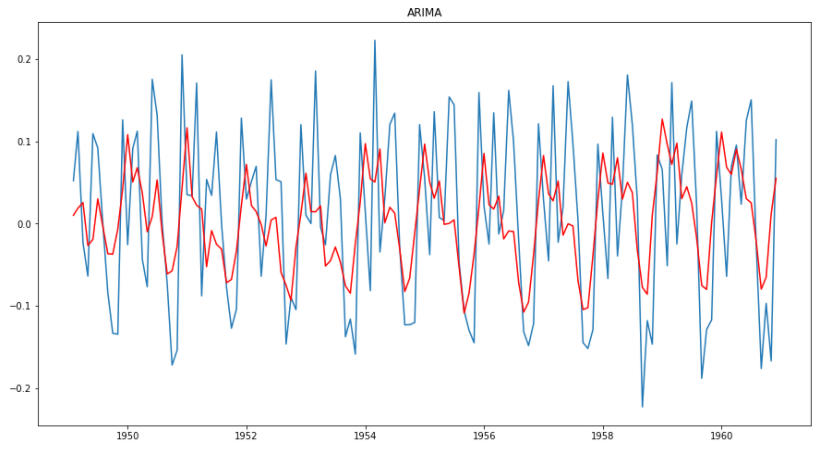
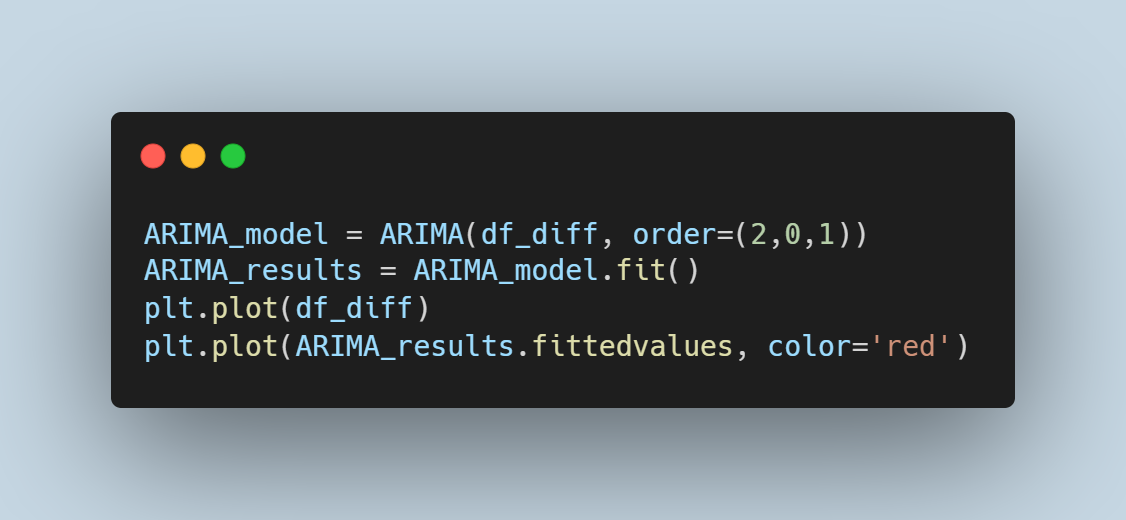
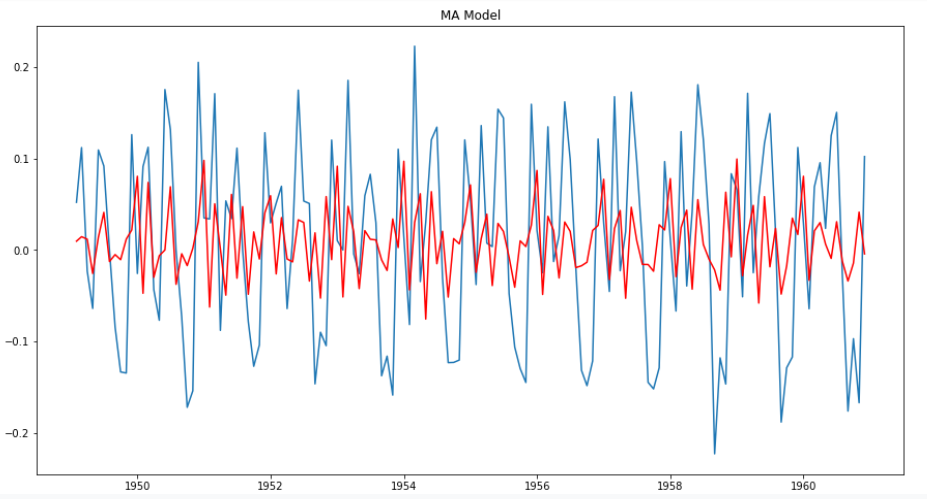
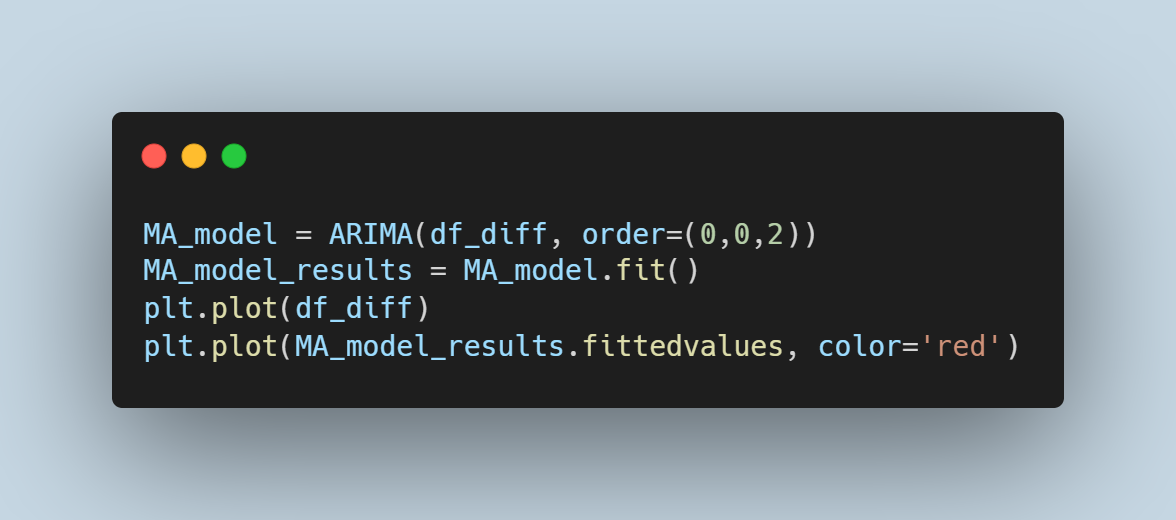
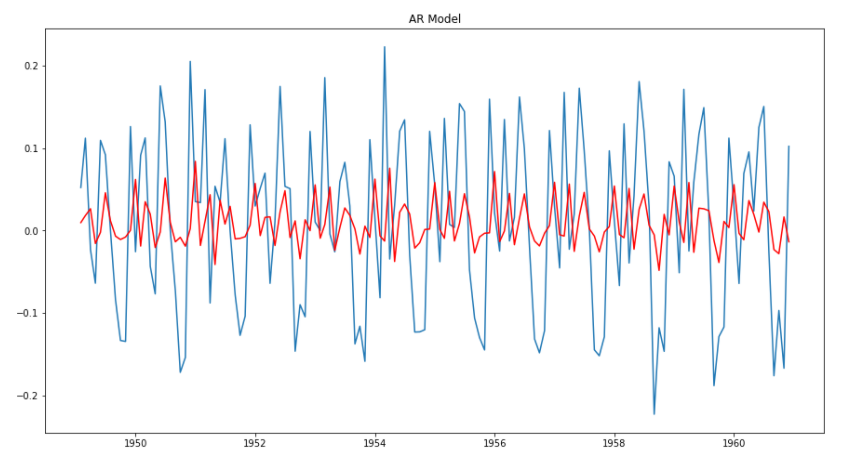
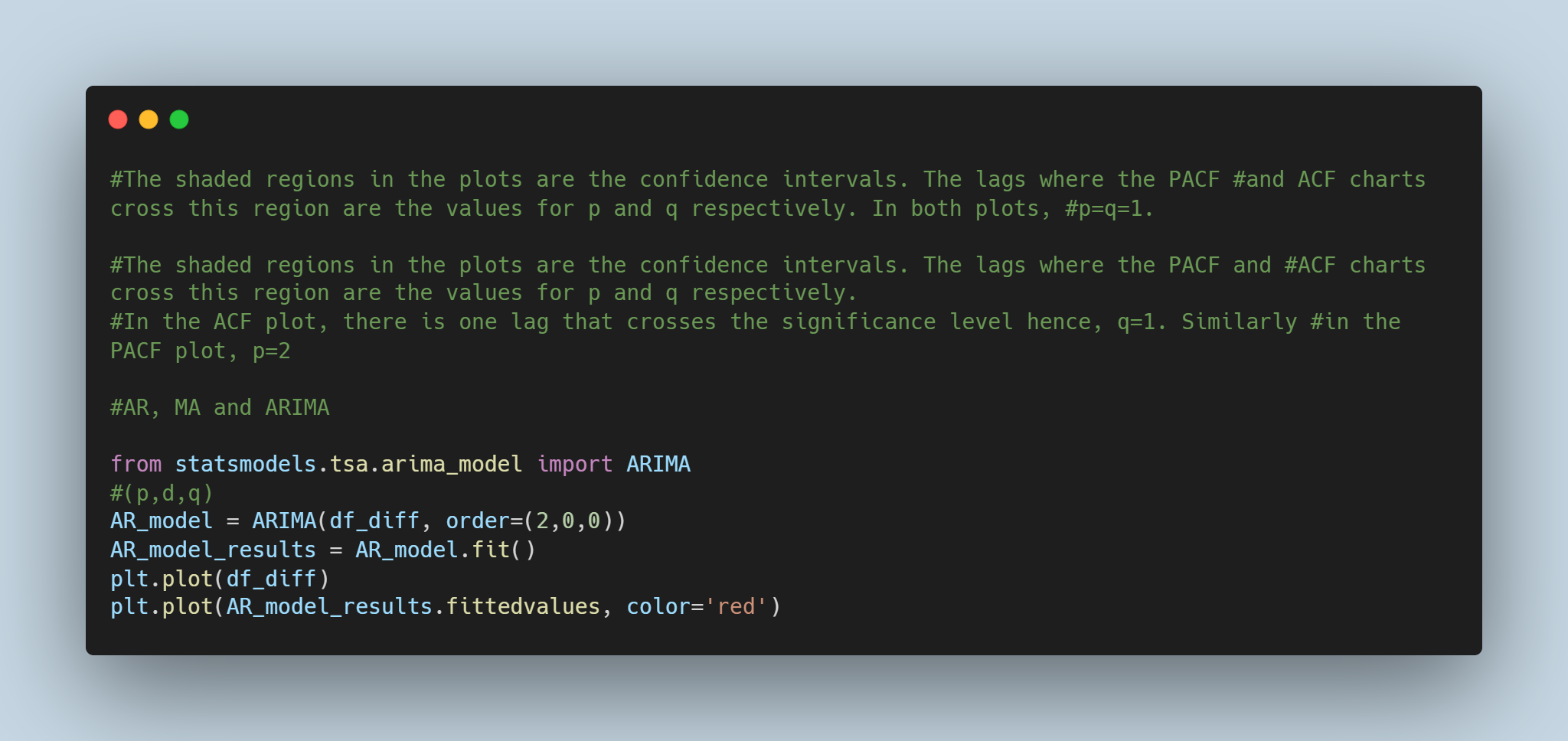
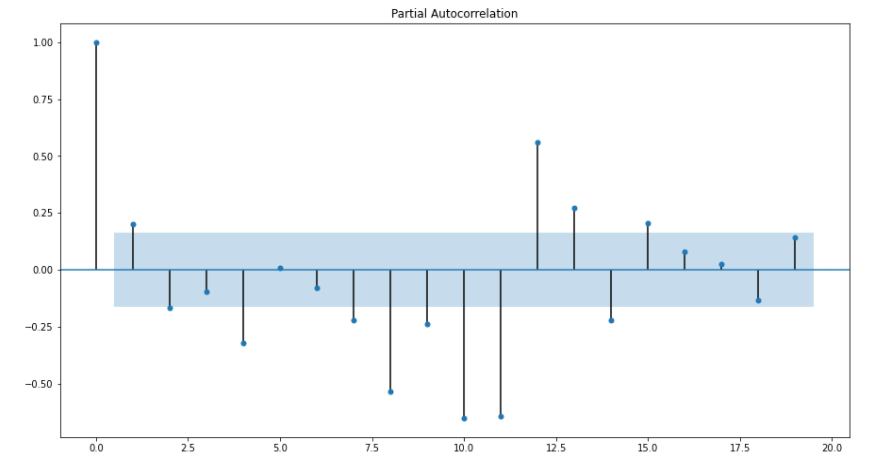
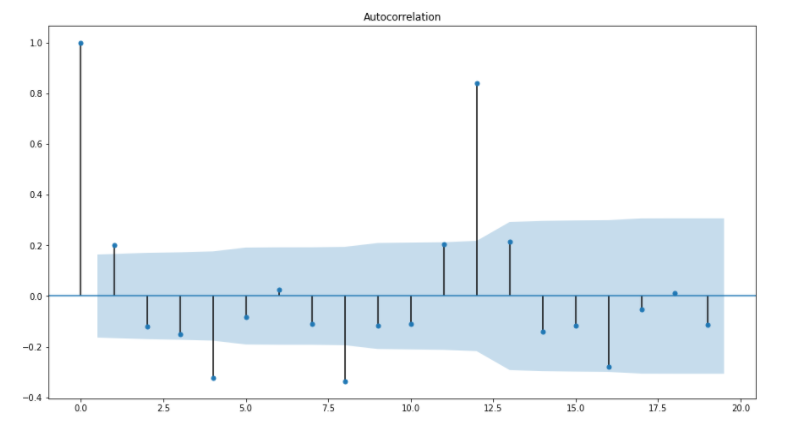
**Autocorrelation Function (ACF):**

Autocorrelation refers to the correlation between the past values in a time series. It measures the linear relationship between lagged values in the series. The ACF plot or correlogram is used to display the autocorrelation in a time series by lag. The correlation coefficient between -1 and 1 is displayed on the y-axis and the lags displayed on the x-axis.

**Partial Autocorrelation Function (PACF):**

PACF gives the partial correlation of a stationary series with its lagged values not accounted for by prior lagged observations. This gives an overview of the relationship between an observation in a series and observations in previous time steps while removing interfering observations. The PACF at a specific lag is the correlation that occurs after removing the effect of any correlations as a result of shorter lags. In contrast to ACF, it controls for other lags.

The ACF and PACF plots are displayed as bar charts showing confidence intervals at 95% and 99% as horizontal lines across the bars. Any bar that crosses these intervals are deemed to be more significant and worth noting.



**Box-Jenkins Methodology**

This is a method  by George Box and Gwilym Jenkins that suggests that the process from which a time series was generated can be modelled using ARMA if stationary and ARIMA if non-stationary. It is an iterative approach that involves a three stage modelling approach which includes identification, estimation and diagnostic checking.

* Identification: The initial step involves identifying if the series is stationary and the number of differences required to make it stationary if it is non-stationary and, to identify the order of the autoregressive and moving average terms - p and q respectively. They recommend a differencing approach to achieve stationarity. Unit roots statistical tests can be used to test for stationarity after each differencing round. It is important to prevent over differencing to avoid including more complexity and extra serial correlation.  ACF and PACF plots are used to obtain q and p such that if there is a strong cut-off in the PACF after a lag and the ACF trails off after a lag, the model is AR. If the autocorrelations are cut off after a few lags then the last lag with a large value will be the estimated value of q. However an autoregressive model or an ARIMA model with positive values for p and q are obtained if the autocorrelations do not cut off. For a MA model, there is a strong cut-off in the ACF after the lag and the PACF trails off after a lag. The PACF determines p such that if the partial autocorrelations are cut off after a few lags, p will be the last lag with a large value. Similarly, if the partial autocorrelations do not cut off, a moving average model is obtained or an ARIMA model with positive p and q values.
* Estimation: Numerical methods that involve maximum likelihood estimation is used to minimize the error term. This can be a calculation intensive and complicated process the reason why the parameter estimation is left to be solved by software.
* Diagnostic checking: The model is checked properly to ensure that it is not overfitting and capturing random noise. An ideal model should have errors that resemble white noise. In addition, the residuals of the model are inspected to confirm that there is no additional structure. Finally, a suitable model is selected.