Forecasting

Commonly used models:

Seasonal indexing

Autoregressive

Moving average

Autoregressive moving average

Autoregressive integrated moving average

Seasonal Indexing

It is used when we have to predict values based on seasonal data. Seasonality is a pattern which occurs within a year.

Steps to forecast the value using seasonal indexing:

Step 1-> calculate the average number of passengers travelling in each year

Step 2-> divide the monthly data by the average of the corresponding year.

Step 3-> compute the average across each year for the values of each month from step 2 in order to obtain the seasonal index

The seasonal index SI can be used to calculate the increase or decrease in predicted values using formula:

1-SI \* 100

Step 4-> divide the data by the seasonal index to obtain the data that does not have a seasonal factor

Step 5-> use linear regression to forecast the trend values for each month

Step 6-> calculate the monthly occupancy for year 1960 by multiplying seasonality index with trend value

Sum of seasonal index is always equal to the seasonal span. For yearly data it is 12, for quarterly data it is 4, monthly it is 1.

Mean of seasonal index is always 1

The autoregressive and moving average models are used as stand alone models as well as in a combination depending on the type of data we are working with.

In order to use autoregressive, moving average or autoregressive moving average models, we need to ensure that the series is stationary. Recall that we have learnt the properties of a stationary series:

* mean is constant
* variance is constant
* co-variance is constant

Now, let us understand how to convert a series in to a stationary series and also to test the stationarity of a series.

Auto Regressive (AR) Model

Autoregressive model works only on stationary data and is used to examine the relationship between different values of the same variable. It is a linear model that can be used to predict future values based on past and present values.

It is represented as AR(p), where p denotes the order or the number of observations considered.

AR(p) :  **y*(t)* =** **β(*0)* + β(*1)*y*(t-1)* + β(*2)*y*(t-2)* + β(3)y*(t-3)* + … + β*(p)*y*(t-p)*+ ε*(t)***

Where 'β' is the coefficient & 'ε' is an error.

For p=1 and p=3

AR(1)  :  y*(t)* = β*0* + β*1*y*(t-1)*+ ε*(t)*

AR(3)  :  y*(t)* = β*0* + β*1*y*(t-1)* + β*2*y*(t-2)* + β*3*y*(t-3)*+ ε*(t)*

**Conclusion**

Autoregressive model helps us to predict the future. There are many ways to identify an appropriate order of the model. Partial auto correlation function (PACF) plot is one of them, which we will see later in this course. It is also possible to predict the future values by using the error terms i.e., Moving average.

Moving Average (MA) Model

The moving average model uses the past errors that result from the regression model to predict the future values.

The future values are calculated by multiplying the past errors with their corresponding coefficients.

Here ε*(t)* represents the error at time 't'. It is an independent and identically distributed data.

MA(q) : **y*(t)* = β*(0)* + α*(0)*ε*(t)* + α*(1)*ε*(t-1)* + α*(2)*ε*(t-2)* + α*(3)*ε*(t-3)*+ … + α*(q)*ε*(t-q)***

Here, q denotes the order of the moving average.

Moving average is represented as MA(q), where q is the order of the model, which indicates how many previous errors we consider to predict the present data

MA(1)  :  y*(t)* = β*0* + α*0*ε*(t)* + α*1*ε*(t-1)*

MA(3)  :  y*(t)* = β*0* + α*0*ε*(t)* + α*1*ε*(t-1)* + α*2*ε*(t-2)* + α*3*ε*(t-3)*

Where 'β0' and 'α' are the coefficient & 'ε' is a error term with mean zero and a constant variance.

We can determine the order of the moving average model by observing the auto correlation function(ACF) plot, which we will see in the later part of this course. We might also use a combination of autoregressive and moving average for model building which will lead us to more accurate prediction, but it all depends on the data.

ARIMA Model

Autoregressive Integrated Moving Average which is also know as Box-Jenkins models which may include autoregressive, moving average and differencing.

It is referred as AR if it uses only autoregressive model, MA if it uses Moving average. Differencing order refers to successive first difference.

 It is used for forecasting trend and assumes that data are correlated with the past data values.

* Non-seasonal ARIMA(p,d,q)
  + p: Autoregressive Order
  + d: Integration Order
  + q: Moving Average Order
* Seasonal ARIMA(p,d,q)x(P,D,Q)s
  + P=number of seasonal autoregressive (SAR) terms,
  + D=number of seasonal differences,
  + Q=number of seasonal moving average (SMA) terms
  + S = seasonality period

Here p,d,q are the non-seasonal terms and P,D,Q are the seasonal terms.

When both p and q are not equal to 0 and d=0, then a combination of AR and MA can be used. This is known as the ARMA model.

* ARMA(p,q)
  + p: Autoregressive Order
  + q: Moving Average Order

The model is referred as:

**y(t) =** **AR(p) + MA(q)**

**y(t) =** **β(0) +** (**β(1)y(t-1) + β(2)y(t-2) + β(3)y(t-3) + … + β(p)y(t-p)) + (α(0)ε(t) + α(1)ε(t-1) + α(2)ε(t-2) + α(3)ε(t-3)+ … + α(q)ε(t-q))**

For ARMA(1,1) and ARMA(3,3):

ARMA(1,1):  y(t) = β(0) + (β(1)y(t-1)) + (α(0)ε(t) + α(1)ε(t-1))

ARMA(3,3):  y(t) = β(0) + (β(1)y(t-1) + β(2)y(t-2) + β(3)y(t-3)) + (α(0)ε(t) + α(1)ε(t-1) + α(2)ε(t-2) + α(3)ε(t-3))

p and q can be determined by observing the auto correlation and partial auto correlation plot.

Different models can be obtained for various combinations of AR and MA individually and collectively.

The best model is obtained by following the diagnostic testing procedure. Below are the two measures for the goodness of fit. The measure trade-off between model fit and complexity of the model.

**1. Akaike Information Criterion(AIC)**

AIC = -2ln(L) + 2k

where L is the value of likelihood function

k is the number of estimated parameters

**2. Bayesian Information Criterion(BIC)**

BIC = -2ln(L) + ln(N)k

where L is the value of likelihood function

N is the number of observations

k is the number of estimated parameters

# Conclusion

The model with the lowest value of the above criterion(AIC and BIC) is chosen as the best model.

ACF and PACF

Before we consider the other two models: ARMA and ARIMA, we need to evaluate the values of p,d and q. This will help us select the appropriate ARIMA(p,d,q) model.

The autocorrelation function(ACF) and partial autocorrelation function(PACF) are used to estimate the parameters of ARIMA(p,d,q).

* Autocorrelation function is a simple correlation between current observation (Yk) and the observation p periods from the current one (Y(t-k))
* Partial autocorrelations are used to measure the degree of association between Yt and Y(t-k) when the effects at other time lags 1,2,3,.., (k-1) are removed.