Time Series - Forecasting

1. Time sries?

* A time-series data is a series of data points or observations recorded at different or regular time intervals. In general, a time series is a sequence of data points taken at equally spaced time intervals. The frequency of recorded data points may be hourly, daily, weekly, monthly, quarterly or annually.

1. There are various terms and concepts in time series that we should know. These are as follows:-

* Dependence- It refers to the association of two observations of the same variable at prior time periods.
* Stationarity - It shows the mean value of the series that remains constant over the time period. If past effects accumulate and the values increase towards infinity then stationarity is not met.
* Differencing- Differencing is used to make the series stationary and to control the auto-correlations. There may be some cases in time series analyses where we do not require differencing and over-differenced series can produce wrong estimates.
* Specification - It may involve the testing of the linear or non-linear relationships of dependent variables by using time series models such as ARIMA models.
* Exponential Smoothing - Exponential smoothing in time series analysis predicts the one next period value based on the past and current value. It involves averaging of data such that the non-systematic components of each individual case or observation cancel out each other. The exponential smoothing method is used to predict the short term prediction.
* Curve fitting - Curve fitting regression in time series analysis is used when data is in a non-linear relationship.
* ARIMA - ARIMA stands for Auto Regressive Integrated Moving Average
* SARIMA - Seasonal Auto Regressive Integrated Moving Average

1. What is forecasting?

* Forecasting is a common statistical task in business, where it helps to inform decisions about the scheduling of production, transportation and personnel, and provides a guide to long-term strategic planning.
* Forecasting a time series can be broadly divided into two types:
* If we use only the previous values of the time series to predict its future values, it is called Univariate Time Series Forecasting.
* If we use predictors other than the series (like exogenous variables) to forecast it is called Multi Variate Time Series Forecasting.

1. Determining what to forecast?

* In the early stages of a forecasting project, decisions need to be made about what should be forecast. For example, if forecasts are required for items in a manufacturing environment, it is necessary to ask whether forecasts are needed for:
* every product line, or for groups of products?
* every sales outlet, or for outlets grouped by region, or only for total sales?
* weekly data, monthly data or annual data?

1. Time plots?

* For time series data, the obvious graph to start with is a time plot. That is, the observations are plotted against the time of observation, with consecutive observations joined by straight lines.

1. Time series patterns?

* Any time series visualization may consist of the following components:

Base Level + Trend + Seasonality + Error.

* Trend: A trend exists when there is a long-term increase or decrease in the data. It does not have to be linear.
* Seasonal: A seasonality is observed when there is a distinct repeated pattern observed between regular intervals due to seasonal factors. A seasonal pattern occurs when a time series is affected by seasonal factors such as the time of the year or the day of the week. Seasonality is always of a fixed and known frequency
* Cyclic: A cycle occurs when the data exhibit rises and falls that are not of a fixed frequency.

1. Stationary and Non-Stationary Time Series?

* Stationarity is a property of a time series. A stationary series is one where the values of the series is not a function of time. So, the values are independent of time.Hence the statistical properties of the series like mean, variance and autocorrelation are constant over time. Autocorrelation of the series is nothing but the correlation of the series with its previous values.A stationary time series is independent of seasonal effects as well.
* We can covert any non-stationary time series into a stationary one by applying a suitable transformation. Mostly statistical forecasting methods are designed to work on a stationary time series. The first step in the forecasting process is typically to do some transformation to convert a non-stationary series to stationary.

1. How to make a time series stationary?

* We can apply some sort of transformation to make the time-series stationary. These transformation may include:

a. Differencing the Series (once or more)

b. Take the log of the series

c. Take the nth root of the series

d. Combination of the above

The most commonly used and convenient method to stationarize the series is by differencing the series at least once until it becomes approximately stationary.

1. differencing:

* If Y\_t is the value at time t, then the first difference of Y = Yt – Yt-1. In simpler terms, differencing the series is nothing but subtracting the next value by the current value.
* If the first difference doesn’t make a series stationary, we can go for the second differencing and so on.

For example, consider the following series: [1, 5, 2, 12, 20]

First differencing gives: [5-1, 2-5, 12-2, 20-12] = [4, -3, 10, 8]

Second differencing gives: [-3-4, -10-3, 8-10] = [-7, -13, -2]

9. Reasons to convert a non-stationary series into stationary one before forecasting?

There are reasons why we want to convert a non-stationary series into a stationary one. These are given below:

* Forecasting a stationary series is relatively easy and the forecasts are more reliable.
* An important reason is, autoregressive forecasting models are essentially linear regression models that utilize the lag(s) of the series itself as predictors.
* We know that linear regression works best if the predictors (X variables) are not correlated against each other. So, stationarizing the series solves this problem since it removes any persistent autocorrelation, thereby making the predictors(lags of the series) in the forecasting models nearly independent.

10. How to test for stationarity?

* This can be done using statistical tests called Unit Root Tests. This test checks if a time series is non-stationary and possess a unit root.

There are multiple implementations of Unit Root tests like:

a. Augmented Dickey Fuller test (ADF Test)

b. Kwiatkowski-Phillips-Schmidt-Shin – KPSS test (trend stationary)

c. Philips Perron test (PP Test)

1. Augmented Dickey Fuller test or (ADF Test) is the most commonly used test to detect stationarity. Here, we assume that the null hypothesis is the time series possesses a unit root and is non-stationary. Then, we collect evidence to support or reject the null hypothesis. So, if we find that the p-value in ADF test is less than the significance level (0.05), we reject the null hypothesis.
2. Kwiatkowski-Phillips-Schmidt-Shin – KPSS test (trend stationary) The KPSS test, on the other hand, is used to test for trend stationarity. The null hypothesis and the P-Value interpretation is just the opposite of ADH test.
3. The Philips Perron or PP test is a unit root test. It is used in the time series analysis to test the null hypothesis that a time series is integrated of order 1. It is built on the ADF test discussed above.

11. How to test for seasonality of a time series?

* The common way to test for seasonality of a time series is to plot the series and check for repeatable patterns in fixed time intervals. So, the types of seasonality is determined by the:-clock or the calendar, Hour of day, Day of month, Weekly, Monthly, Yearly

However, if we want a more definitive inspection of the seasonality, use the Autocorrelation Function (ACF) plot. There is a strong seasonal pattern, the ACF plot usually reveals definitive repeated spikes at the multiples of the seasonal window.

12. ACF & PACF?

* ACF(auto-correlation function) : It describes how well the present value of the series is related to its past values.
* PACF(partial autocorrelation function) : Instead of finding correlations of present with lags like ACF, it finds the correlations of the residuals with the next lag value thus ‘partial’ and not ‘complete’ as we remove already found variations before we find next correlation.

13. simple forecasting methods?

* Avg method: Here, the forecasts of all future values are equal to the average (or “mean”) of the historical data.
* Naive method: For naïve forecasts, we simply set all forecasts to be the value of the last observation. This method works remarkably well for many economic and financial time series.

14. Time series decomposition?

1. Moving avg method:

* This model states that the next observation is the mean of all past observations.
* Assigns equal weights to all past observations
* Better to forecasrt when data & environment is not volatile
* Window width is key to success

1. Simple exponential smoothing:

* The simplest of the exponentially smoothing methods is naturally called simple exponential smoothing (SES). This method is suitable for forecasting data with no clear trend or seasonal pattern
* Assigns more weight to recent obs than past obs
* Smoothing constant value is key to success

15. Trend methods?

1. Holt’s linear trend method: Holt (1957) extended simple exponential smoothing to allow the forecasting of data with a trend.

* This method involves a forecast equation and two smoothing equations (one for the level and one for the trend)

1. Holt-Winters’ seasonal method:

* Holt (1957) and Winters (1960) extended Holt’s method to capture seasonality. The Holt-Winters seasonal method comprises the forecast equation and three smoothing equations
* Depending on the nature of the trends and seasonality, a time series can be modeled as an additive or multiplicative time series. Each observation in the series can be expressed as either a sum or a product of the components.

1. The additive method is preferred when the seasonal variations are roughly constant through the series

* Additive time series: Value = Base Level + Trend + Seasonality + Error

1. While the multiplicative method is preferred when the seasonal variations are changing proportional to the level of the series.

* Multiplicative Time Series: Value = Base Level x Trend x Seasonality x Error

16. What is ARIMA & its models?

* ARIMA stands for Autoregressive Integrated Moving Average Model. It belongs to a class of models that explains a given time series based on its own past values -i.e.- its own lags and the lagged forecast errors. The equation can be used to forecast future values. Any ‘non-seasonal’ time series that exhibits patterns and is not a random white noise can be modeled with ARIMA models.
* So, ARIMA, short for Auto Regressive Integrated Moving Average, is a forecasting algorithm based on the idea that the information in the past values of the time series can alone be used to predict the future values.

ARIMA( p,d,q) model, where

p = order of the autoregressive part;

d = degree of first differencing involved;

q = order of the moving average part.

* AR(p) Autoregression – a regression model that utilizes the dependent relationship between a current observation and observations over a previous period.
* I(d) Integration – uses differencing of observations (subtracting an observation from observation at the previous time step) in order to make the time series stationary. Differencing involves the subtraction of the current values of a series with its previous values d number of times.
* MA(q) Moving Average – a model that uses the dependency between an observation and a residual error from a moving average model applied to lagged observations. A moving average component depicts the error of the model as a combination of previous error terms. The order q represents the number of terms to be included in the model.
* Types of ARIMA Model:

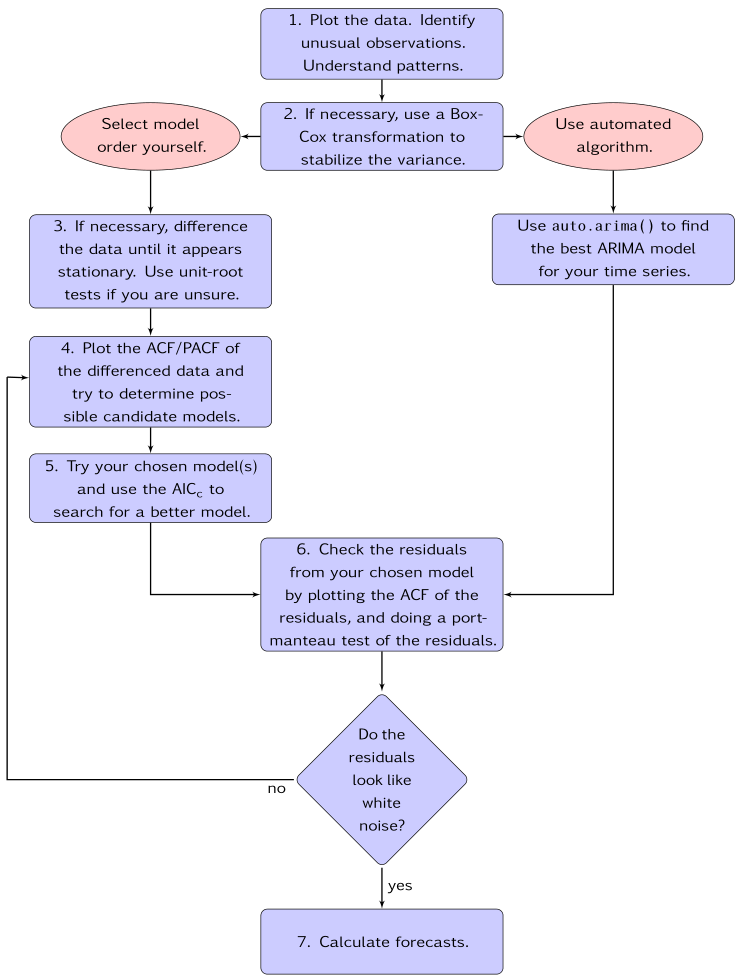
ARIMA : Non-seasonal Autoregressive Integrated Moving Averages

SARIMA : Seasonal ARIMA

SARIMAX : Seasonal ARIMA with exogenous variables

If a time series , has seasonal patterns, then we need to add seasonal terms and it becomes SARIMA, short for Seasonal ARIMA.

ARIMA MODEL FLOW CHART



17. How to find the order of differencing (d) in ARIMA model ?

* As stated earlier, the purpose of differencing is to make the time series stationary. But we should be careful to not over-difference the series. An over differenced series may still be stationary, which in turn will affect the model parameters.
* So we should determine the right order of differencing. The right order of differencing is the minimum differencing required to get a near-stationary series which roams around a defined mean and the ACF plot reaches to zero fairly quick.
* If the autocorrelations are positive for many number of lags (10 or more), then the series needs further differencing. On the other hand, if the lag 1 autocorrelation itself is too negative, then the series is probably over-differenced.
* If we can’t really decide between two orders of differencing, then we go with the order that gives the least standard deviation in the differenced series.

Now, lets understand the concepts with the help of an example as follows:-

First, I will check if the series is stationary using the Augmented Dickey Fuller test (ADF Test), from the statsmodels package. The reason being is that we need differencing only if the series is non-stationary. Else, no differencing is needed, that is, d=0.

The null hypothesis (Ho) of the ADF test is that the time series is non-stationary. So, if the p-value of the test is less than the significance level (0.05) then we reject the null hypothesis and infer that the time series is indeed stationary.

So, in our case, if P Value > 0.05 we go ahead with finding the order of differencing.

17. How to find the order of the AR term (p)?

* The next step is to identify if the model needs any AR terms. We will find out the required number of AR terms by inspecting the Partial Autocorrelation (PACF) plot.
* Partial autocorrelation can be imagined as the correlation between the series and its lag, after excluding the contributions from the intermediate lags. So, PACF sort of conveys the pure correlation between a lag and the series. This way, we will know if that lag is needed in the AR term or not.
* Partial autocorrelation of lag (k) of a series is the coefficient of that lag in the autoregression equation of Y.

Yt=α0+α1Yt−1+α2Yt−2+α3Yt−3

* That is, suppose, if Y\_t is the current series and Y\_t-1 is the lag 1 of Y, then the partial autocorrelation of lag 3 (Y\_t-3) is the coefficient α3 of Y\_t-3 in the above equation.
* Now, we should find the number of AR terms. Any autocorrelation in a stationarized series can be rectified by adding enough AR terms. So, we initially take the order of AR term to be equal to as many lags that crosses the significance limit in the PACF plot.

18. How to find the order of the MA term (q)

* Just like how we looked at the PACF plot for the number of AR terms, we will look at the ACF plot for the number of MA terms. An MA term is technically, the error of the lagged forecast.
* The ACF tells how many MA terms are required to remove any autocorrelation in the stationarized series.

19. Accuracy Metrics for Time Series Forecast

* The commonly used accuracy metrics to judge forecasts are:
* Mean Absolute Percentage Error (MAPE)
* Mean Error (ME)
* Mean Absolute Error (MAE)
* Mean Percentage Error (MPE)
* Root Mean Squared Error (RMSE)
* Lag 1 Autocorrelation of Error (ACF1)
* Correlation between the Actual and the Forecast (corr)
* Min-Max Error (minmax)
* Typically, we will use three accuracy metrices:-
* MAPE
* Correlation and
* Min-Max Error can be used.

The above three are percentage errors that vary between 0 and 1. That way, we can judge how good the forecast irrespective of the scale of the series is.

20. SARIMA model in python?

* The plain ARIMA model has a problem. It does not support seasonality.
* If the time series has defined seasonality, then we should go for Seasonal ARIMA model (in short SARIMA) which uses seasonal differencing.
* Seasonal differencing is similar to regular differencing, but, instead of subtracting consecutive terms, we subtract the value from previous season.
* So, the model will be represented as SARIMA(p,d,q)x(P,D,Q), where, P, D and Q are SAR, order of seasonal differencing and SMA terms respectively and 'x' is the frequency of the time series. If the model has well defined seasonal patterns, then enforce D=1 for a given frequency ‘x’.
* We should set the model parameters such that D never exceeds one. And the total differencing ‘d + D’ never exceeds 2. We should try to keep only either SAR or SMA terms if the model has seasonal components.

20. SARIMAX model with exogenous variables?

* Now, we will force an external predictor, also called, exogenous variable into the model. This model is called the SARIMAX model. The only requirement to use an exogenous variable is we should know the value of the variable during the forecast period as well.

21. Deep Learning

Prediction of number of passengers for an airline using LSTM

In this project we are going to build a model to predict the number of passengers in an airline. To do so we are going to use Recurrent Neural Networks, more precisely Long Short Term Memory.

Recurrent Neural Network

* Neural Networks are set of algorithms which closely resemble the human brain and are designed to recognize patterns.
* Recurrent Neural Network is a generalization of feed forward neural network that has an internal memory.
* RNN is recurrent in nature as it performs the same function for every input of data while the output of the current input depends on the past one computation.
* After producing the output, it is copied and sent back into the recurrent network. For making a decision, it considers the current input and the output that it has learned from the previous input.
* In other neural networks, all the inputs are independent of each other. But in RNN, all the inputs are related to each other.

Long Short Term Memory

* Long Short-Term Memory (LSTM) networks are a modified version of recurrent neural networks, which makes it easier to remember past data in memory.
* Generally LSTM is composed of a cell (the memory part of the LSTM unit) and three “regulators”, usually called gates, of the flow of information inside the LSTM unit: an input gate, an output gate and a forget gate.
* Intuitively, the cell is responsible for keeping track of the dependencies between the elements in the input sequence.
* The input gate controls the extent to which a new value flows into the cell, the forget gate controls the extent to which a value remains in the cell and the output gate controls the extent to which the value in the cell is used to compute the output activation of the LSTM unit.
* The activation function of the LSTM gates is often the logistic sigmoid function.
* There are connections into and out of the LSTM gates, a few of which are recurrent. The weights of these connections, which need to be learned during training, determine how the gates operate.

