Now Splitting the data into train and split we cannot randomly split the data into two part we have to keep the sequence of the time series intact in order to use it for forecasting after a lot of trial and error we found 20 years or less than this is the optimal split for our time series. we have kept 20 years of data in testing and the rest of the data in training.

Thanks swapnil. Hello everyone good day to you all. I am Dalvi Moin let’s continue from here.

Note: our data has irregularity, it doesn’t have any seasonality

### It has no gradual upward trend but it is not linear trend.it is polynomial trend

### The data is affected by unexpected situations that may lead in spikes in a short amount of time span.

There are Random variations and fluctuations. The time series do not have any definite pattern. Thus, these variations are not purely with respect to time but may be due other factor that influence to increase or decrease in emission of CO2 which could world wars, deforestation as well as the burning of fossil fuels like coal, oil and natural gas., etc. It is difficult to forecast these kind of time series

Now model building. We have built the model on two different data set

Obviously Training data is used for model building and testing data is used for evaluation of the model.

1. Resampled data set in which we converted yearly data into monthly data set using Linear interpolation
2. Linear Interpolation simply means to we are estimating the missing value by connecting a straight line between the previous known value and the next known value
3. In short, in our case to convert from yearly to monthly first we add the month and take January as known and the next years January as known connecting these dots and make straight line.
4. We are going to compare both the results of model building on these two dataset and not only select the model but also the best dataset for forecasting.

Next slide

**Our test data doesn’t have linear upward trend there is a downfall after the 2005 and before there till 2000 there is linear upward trend**

Next slide

In this plot we are comparing the forecasted or predicted values of different models with respect to our test data

The blue line here is our test data and the rest are our models predictions as we can visually see none of the models came close to the test data. They have predicted trend in upward linear manner but no one could predict the down fall in the test data.

These are holt’s winter methods simple exponential double and triple exponential.

Triple Exponential considers the trend and seasonality in our data set

Double exponential doesn’t considers the seasonality rather it considers there is a trend in our dataset

And simple exponential dataset doesn’t considers the trend or seasonality in our dataset

simple exponential only considers there is a constant trend in dataset.

All these methods are trying to capture a linear trend but our data doesn’t have a linear trend.

Next slide

Here is the plot of all the model evaluated on the resampled dataset we have considered two metrics for evaluation

1. RMSE (Root mean squared Error)

Let me break it down for you let take Error it is the difference between the actual value (test data) and the forecasted value. Next is Squared why we need to square because the error value we get not only positive but also be negative as well to remove the negative part in error we can either square or take the absolute value . in this case we choose square of the error now the next is Mean it is to calculate the average amount of error we are getting through our model for the complete test data. Mean is also the value where most of our data lies in our data set. After squaring the error we now have changed the scale of the data set to make it comparable we need to revert the effect of square to do that we have Square Root option.

1. Another Method does the same as RMSE but in a different way and the end result is in the form of percentage . The percent of error or the Error rate we are getting for our model.

AS we can see the Simple Exponential did better than the other models with 5% of error rate and 1.06 of rmse

Next slide

After forecasting the value and plotting with and comparing the values we can see the best model is not even doing it best it is just predicting the last value and giving us a constant line with no trend .

Next slide

In this plot we are comparing the training and testing error

WE can see here as compared to the training error we are getting a significant amount of testing error

Next slide

Now we are going to build the same models but on the raw data set which had yearly time series

As we can see in this plot most of the model has captured the linear trend and some variation seasonality . but none of them captured the downfall. They are forecasting a linear upward trend with some seasonality

Next slide

Here we have tried the ARIMA model with hyperparameter with respect to the minimum RMSE resultant. We are going to select those combination which gives us the least RMSE

In this plot we can observe it has

Now let see the results for these model in our

Next slide