# Week 11
# Artificial Intelligence (AI) for Investments

# Lesson 2: Regression Algorithm: Application

# Introduction

- Application of regression algorithm in prediction of security prices
- ABC case study
- Simple linear regression
- Multiple linear regression
- Summary and concluding remarks

# Case Study: ABC Stock Price

# Case Study: Sentiment Problem

- Stock price prediction or stock return prediction is an attempt to determine the future value of a company based on analysis of factors, which impact its price movement
- There are a number of factors that help in predicting stock prices
- These can be macroeconomic factors like state of the country's economy, growth rate inflation, etc.
- There are also other factors that are more specific to a stock like profit margin, debt to equity issues, sales of a company, and so on

# Case Study: Sentiment Problem

So we are given the data for stock market price for ABC company, along with Nifty and Sensex (market indices). We are also given the data of dividend announcement and a sentiment index

| Date | Price | ABC | Sensex | Dividend Announced | Sentiment | Nifty |
|---|---|---|---|---|---|---|
| 03-01-2000 | 718.15 | 0.079925 | 0.073772 | 0 | 0.048936 | 0.095816 |
| 04-01-2000 | 712.9 | -0.00731 | 0.021562 | 0 | -0.05504 | 0.009706 |
| 05-01-2000 | 730 | 0.023987 | -0.02441 | 0 | 0.019135 | -0.03221 |
| 06-01-2000 | 788.35 | 0.079932 | 0.012046 | 0 | 0.080355 | 0.011205 |
| 07-01-2000 | 851.4 | 0.079977 | -0.0013 | 0 | 0.094038 | -0.0004 |
| 10-01-2000 | 919.5 | 0.079986 | 0.019191 | 1 | 0.015229 | 0.030168 |
| 11-01-2000 | 880 | -0.04296 | -0.04025 | 0 | -0.07217 | -0.04966 |
| 12-01-2000 | 893.75 | 0.015625 | 0.036799 | 0 | 0.01396 | 0.020999 |
| 13-01-2000 | 875 | -0.02098 | -0.00845 | 0 | 0.057518 | -0.01164 |
| 14-01-2000 | 891 | 0.018286 | 0.004858 | 1 | 0.008828 | 0.020714 |
| 17-01-2000 | 819.75 | -0.07997 | -0.01228 | 0 | -0.12395 | -0.00962 |
| ..... | ..... | ..... | ..... | ..... | ..... | ..... |
| ..... | ..... | ..... | ..... | ..... | ..... | ..... |

# Case Study: Sentiment Problem

- Consider a portfolio manager who has built a model for a particular stock
- The manager wants to predict the ABC stock price returns for this stock using regression model
- The data starts from 2007 and goes till 2019, so we have approximately 13 years of data
- We have daily returns of ABC or change in price of ABC in column B. Next, we have daily return on Sensex in column C and daily return on Nifty in column D.

# Case Study: Sentiment Problem

- Sensex and nifty are the two main stock indices used in India
- They are benchmark Indian stock market indices that represent the weighted average of the largest Indian companies
- So, Sensex represent average of 30 largest and most actively traded Indian companies
- Similarly, Nifty represents a weighted average of 50 largest Indian companies.

# Case Study: Sentiment Problem

- Another variable is dividend announcement in column E, which is one, if a company has announced dividend on a particular date and zero otherwise
- So, for example, it is one on January 2, 2007, because the company ABC announced a dividend on this date and it is zero for all other days when the company did not announce any dividend. Notice that this is a dummy variable

# Case Study: Sentiment Problem

- Lastly, we have a sentiment variable in column F. It is a sentiment score which quantifies how investors feel about ABC
- It can be based upon news analysis or upon option market analysis or based on some survey
- We would not go into details of the score here and take it as given. A very high sentiment score represents bullish investors and vice versa.

# Case Study: Problem Statement

# Case Study: Problem Statement

The following tasks need to be performed: Part 1

- Data Visualization
- Training the model
- Testing the model
- Evaluate out-of-sample performance of the model

# Case Study: Problem Statement

The following tasks need to be performed: Part 2

- Training and testing the model using multiple linear regression algorithm
- Testing the model
- Examine issues in estimation and how to resolve them
- Evaluate out-of-sample performance of the model

# Case Study: Data Input

# Introduction

We start with the R implementation of the case study problem statement

# Summary

To summarize the video, first we loaded the relevant packages and libraries, then we set the working directory, and finally we read the “ABC” data file in R

In the next video we will try to visualize various properties of the data

# Case Study: Data Input

# Introduction

We start with the R implementation of the case study problem statement

# Summary

To summarize the video, first we will loaded the relevant packages and libraries, then we set the working directory, and finally we read the “ABC” data file in R

In the next video we will try to visualize various properties of the data

# Case Study: Data Visualization

# Introduction

In this video we will examine the key variables in the data through visualization

We will visualize the returns on ABC and Nifty

We will also visualize the cumulative returns for ABC and Nifty

# Summary

To summarize the video, we visualized the returns and cumulative returns for ABC and Nifty returns using R programming

# Case Study: Data Visualization

# Introduction

In this video we will examine the key variables in the data through visualization

We will visualize the returns on ABC and Nifty

We will also visualize the cumulative returns for ABC and Nifty

# Summary

To summarize the video, we visualized the returns and cumulative returns for ABC and Nifty returns using R programming

In the next video, we will examine the summary measures

# Case Study: Data Summary

# Introduction

In this video, we will discuss the basic properties of the data and summary measures

# Summary

To summarize the video, first we summarized the key return variables

Next we plotted the density distribution of these variables

We noted that ABC returns are heavily skewed towards the left

# Case Study: Normality

# Introduction

In this video we will examine the normality of the data

This includes examination of skewness and kurtosis measures

# Summary

To summarize the video, we computed the skewness and kurtosis measures for the data

Data appears to be left skewed

Then we also examined the statistical significance of the skewness, kurtosis measures and also conducted the Jarque-Bera test of normality

# Case Study: Stationarity

# Introduction

In this video, we will discuss the stationarity property and conduct the examination of data stationarity in R

# Summary

To summarize the video, we conducted tests of data stationarity

These included ADF, PP, and KPSS tests

We found that data is stationary

# Case Study: Training and Test Data

# Introduction

In this video, we will segregate the “ABC” data into training and test data

Training data is employed to train the linear regression algorithm

Test data is employed to test the out-of-sample forecasting efficiency of the algorithm

# Summary

To summarize the video, we segregated our data in two segments

The training data included observations from the year 01-Jan-2007 to 01-Dec-2017, comprising 2850 observations

The test data included observations from the year 04-Jan-2017 onwards, comprising 478 observations

# Training the Simple Linear Regression (SLR) Algorithm

# Introduction

In this video, we will train a simple linear regression algorithm by regressing ABC returns on Nifty returns

# Summary

To summarize the video, we examined the relationship between ABC returns and Nifty returns

To this end, we trained a simple linear regression algorithm

We also reviewed the output of the regression model; we found a significant coefficient for Nifty

We also noted that the model explains around 10.87% variation in ABC returns

# Training the model: Residual Diagnostics

# Introduction

In this video, we will perform the residual diagnostics of the simple linear regression model build using training dataset

# Summary

To summarize the video, we conducted residual diagnostics of the trained model

First, we plotted the density plot of the raw residuals and studentized residuals

Next, we checked the normality of the residuals with the help of qqplot

We also conducted the outlier test

We found certain outliers through these methods; these outliers can be removed to improve the model estimates

# Training the model: Heteroscedasticity

# Introduction

In this video, we will examine the econometric issue of heteroscedasticity or non-constant variance of error terms that afflicts the estimation

# Summary

To summarize the video, first we visualized the issue of heteroscedasticity by plotting the residuals with fitted values

Residuals appeared to have non-constant variance

We conducted the tests of non-constant variance and found that the result is indeed statistically significant

# Training the model: Autocorrelation

# Introduction

In this video, we discuss and empirically test the issue of autocorrelation

# Summary

To summarize the video, we conducted Durbin-Watson and Breusch-Pagan tests of autocorrelation

We find evidence of serial correlation in error terms at higher order

In practical situation, one can account for such serial correlation by adding the lags of variable that is serially correlated

# Training the model: Robust Standard Errors

# Introduction

In this video, we discuss robust standard error to resolve the issue of heteroscedasticity and autocorrelation

# Summary

To summarize the video, we discussed the application of robust standard errors in correcting for issues such as heteroscedasticity and autocorrelation

We discussed four most prominent available rountines (hccm, vcovHAC, vocvHC, and NeweyWest) for correcting the model standard errors

# Prediction with Simple Linear Regression (SLR) Algorithm

# Introduction

We have trained our simple linear regression algorithm and tested with test data

Now, we will employ our trained algorithm for prediction using test data

# Summary

To summarize the video, we forecasted ABC returns using test data

We visualized the ABC actual and predicted returns

The predicted returns have 43.78% correlation with actual returns, and therefore we conclude that regression algorithm has predicted the returns reasonably accurately

# Out-of-sample forecasting efficiency

# Introduction

A model may perform good on the training data, i.e., in-sample goodness-of-fit measures; however, its true capability is established only if performs well in out-of-sample prediction

In this video, we will perform out-of-sample forecasting and prediction  based on the predicted values and actual values of ABC returns

# Summary

To summarize the video, one needs some cost or error function to compare between competing algorithms

In this video we discussed and reviewed the implementation of various cost/error functions (e.g., MSE, RMSE, MAPE, SMAPE, MSLE, etc.)

# Training the Multiple Linear Regression (MLR) Algorithm

# Introduction

In this video, we will implement the multiple linear regression (MLR) algorithm with variables namely ABC Returns, market returns (Nifty and Sensex), Dividend announcements, and Sentiment

# Summary

To summarize the video, we trained our MLR algorithm using training dataset

Then we reviewed the output of the model

We find that the model may be afflicted by the issue of multicollinearity, which will be resolved in the next video

We also note that the variables namely Market returns, Sentiment, and Dividend announcement appear to be significant

The model explains about 27.84% variation in the ABC returns

# Case Study: Multicollinearity

# Introduction

Independent variables may be correlated resulting in multicollinearity

In this video, we will examine the issue of multicollinearity and find ways to resolve the same

# Case Study: Summary

To summarize the video, we computed the correlations across the dependent variables and found that market proxies (Nifty and Sensex) are highly correlated leading to the issue of multicollinearity

This is also corroborated by the high variance inflation factor (VIF) odf ~2.9 with Nifty and Sensex

So we remove one of the market proxy (Nifty) and again train the model and review the model output

# Prediction with MLR

# Introduction

In this video, using our trained MLR algorithm, we will make predictions about ABC returns

Next, we will compare the predicted and actual returns through visualization

We will also compute the correlations between the predicted and actual returns

# Summary

To summarize the video, we predicted the ABC returns using our trained algorithm with the test data

Plots of actual vs predicted returns suggest that our predicted returns are indeed able to mimic our actual returns

Moreover, our predicted returns exhibit a high correlation of about 57.70%, which indicates a high prediction accuracy

# Summary and Concluding Remarks

# Summary and Concluding Remarks

- ABC stock prices are modelled using simple regression problem with market index variable
- Model is trained using training dataset and various goodness-of-fit measures are examined
- Fitted modelled is examined visually as well
- Model is tested using test dataset and various measures of out of sample fit are examined

# Summary and Concluding Remarks

- Next, a multiple linear regression model is trained using training dataset on multiple variables
- Fitted model is visually examined and also various goodness-of-fit measures are examined
- Model is evaluated on various issues related to multicollinearity, heteroscedasticity, and autocorrelation
- Lastly, model is examined on various parameters for its out of sample fit performance

# Thanks!