

Abstract

A good investment is one that can make more profit for investors. The aim of this research is to identify and quantify factors, which give the best proportions of the speculation potential in this dataset and make inferences about which kinds of stock would provide the best investment strategy. The share price as the dependent variable, in the dataset, is the price of a single share of a number of saleable stocks of companies, derivative or other financial assets. In order to evaluate how the factors influence the share price and their degree of correlation on stock (market) performance, this study analyzed stock statistics and the mathematical method used is the multiple- linear regression model that can predict share price for each company. The multiple linear regression model has been developed through the analysis of data from 102 companies in the processing industry and by using the specific R instruments.

Introduction:

The stock exchange is known to be a complex versatile framework that is hard to foresee because of the extensive number of factors that decide the everyday price fluctuations. The Investment strategy is a term in financial control that refers to particular economic plan, regulations, and procedures meant to aid an investor in making a ramification from some of the investment possibilities, constructing and handling an investment portfolio profitably. In this project by applying the multi-regression method, attempts to study the relationship between a dependent variable and independent variables.

Nevertheless, that is an assumption that the project is making to simplify the model due to the use of the selected regression models. This project's objective is to apply Multilinear-regression model to predict the share price adjustments and Assess diverse models' prosperity by withholding information amid preparing and assessing the precision of these forecasts utilizing known data. This study permits methods for real-world Multilinear-regression demonstrate counting procuring and analyzing even a big data set, and using an assortment of methods and techniques to prepare the program and anticipate potential results.

Methodology:

i. Data Collection:

- a. Download the database in 'xlsx' format or from
- b. Import the data to R

The dataset that was used was collected from given investment Database .the attributes collected are listed below it consists of 102 rows and 16 columns.(all attributes are numerical)

The Data Variable Name	Description
1. Share Price (p)	• The share price [in pence] on a particular date.
2. Avg Vol	• Average volume of share sales over previous three months.
3. EPS(Current)	• Earnings per Share(Current)
4. P/E	• Price to Earnings Ratio (Previous Close/EPS)
5. MktCap	• The Market Capitalisation [£Billions](Last Trade x Current Number of Shares Outstanding)
6. Dividend	• The annual dividend per share of stock.
7. Yield	• The annual dividend per share divided by the previous closing stock price [%]
8. Alpha	• Measures whether a stock has a higher or lower returns than expected considering its risk
9. Beta	• It measures the stock responses to the reference index.
10. Employees	• Number of Employees
11. EPS(2016)	• EPS in 2016
12. EPS(2017)	• EPS in 2017
13. Net Revenue(2016)	• Net Revenue in 2016
14. Net Revenue(2017)	• Net Revenue in 2017
15. Net Profit(2016)	• Net Profit in 2016
16. Net Profit(2017)	• Net Profit in 2017

ii. Data Mining:

- c. Identifying the source information.
- d. Picking the data points that need to be analyzed.
- e. Extracting the relevant information from the data.
- f. Drop missing values
- g. Interpreting and reporting the results.
- h. Outlier Detection&removing: using box plotted to view outliers present in that dataset, and The outliers which are out of the range was removed .

iii. Graphical Data Analysis

Histogram:

Histograms are among the most helpful strategies and enable you to pick up understanding into your information, including conveyance, focal inclination, spread, methodology and outliers.

Boxplot:

Plots are curious for speaking to data around the central tendency, Symmetry, skew and exceptions, but they can cover up a few angles of the information such as Multimodality.

Normal -QQ plot:

Scatterplots plotted for checking normality of residual of the final regression model, and Scatterplots w plotted between the dependent variable of dataset and each independent variable (Predictor) to look for relationship if it is linear, also to see influential cases.

iv. Correlation Analysis

It a analysis includes distinguishing the relation between a dependent variable and at least one autonomous factors.the range of correlation is between -1 and 1. A model of the relationship is conjectured, and gauges of the parameter esteems are utilized to build up an expected regression equation.

v. Regression Analysis

- Linear regression of share price :

The Equation for the multi-linear regression model is

$$y_i = \beta_0 + x_{i1}\beta_1 + \dots + x_{ik}\beta_k + e_i \quad i = 1, 2, \dots, n.$$

Where y is the dependent variable, X_1, X_2 up to X_i are P distinct predictors (independent variables), b_0 is the y -intercept which is obtained when all independent variables values are zero, b_1 up to b_p are coefficients estimates of the regression. IN this project, we have 102 variables to choose from, way outside the realm of the all

- Test the fit of the model (Step-selection)

➤ Forward selection

It was chosen as the best selection for the project is a subset of the indicator factors for the last model. Begin with an invalid model. The invalid model has no indicators, only one block (The mean over Y). Fit p simple linear regression models, each with one of the variables in and the intercept. So basically, you just search through all the single-variable models the best one (the one that results in the lowest residual sum of squares). You pick and fix this one in the model. Proceed until the point that some halting principle is fulfilled, for instance when every staying variable has a p -value over some edge. P

➤ **Backward selection :**

- Dissimilar to forward stepwise choice, it starts with the full least squares show containing all p indicators, and after that iteratively expels the minimum helpful indicator, each one in turn. tart with all variables in the model. Remove the variable with the largest p-value | that is, the variable that is the least statistically significant. The new (p - 1)-variable model is, and the variable with the largest p-value is removed. Continue until a stopping rule is reached. For instance, we may stop when all remaining variables have a significant p-value defined by some significance threshold.

Result:

By applying regression model on R and doing backward step the below equation came out :

$$\text{Share Price (p)} = 0.79482 (\text{'P/E'}) + 30.142(\text{Dividend}) + 3.99649(\text{'EPS(Current)}) - 139.212(\text{Yield})$$

Where R-squared: 0.9001, and Adjusted R-squared: 0.8906, which means how close the data **are** to the fitted regression line the model, and high accuracy .a

The forward-step result, which is the better-fit regression model, is:

$$\text{Share Price (p)} = 30.142(\text{Dividend}) - 139.212(\text{Yield})$$

Where Multiple R-squared: 0.8803, Adjusted R-squared: 0.8767

Conclusion:

In this study, we modeled selected internal factors of selected stock dataset, using multiple linear regression to establish the factors that determine the best investment model for companies. The study aimed to ascertain the relationship between 'share price' and the selected internal factors, to identify the internal factors in share price of companies that affect their investment strategies. In this study several R function used to find best fit regression model for our dataset and at the end ,the result showed that with applying forward step regression model **yield** and **Dividend** are the most important independent factors to predict share price which help companies to have better investment strategy.