# Abstract

Fundamental analysis is an important tool for investors looking for opportunities to invest in the stock market. It is also a useful tool for management who can use it as a tool to identify areas to focus on in order to generate value for shareholders. Changes in fundamentals from one year to another reflect in the stock price changes over the same period of time. Linear regression is shown to be a sufficient model for evaluating the significance of changes in fundamentals on the stock price. Apart from the company’s fundamentals that are updated periodically, the sector within which a company operates also provides relevant information on the effects of external forces that impact on the stock price. Management can use fundamental analysis to facilitate decision making in-line with maximizing shareholder value.

# Introduction

Stock markets provide an opportunity for investors to own a proportion of a company and earn returns in the form of dividends and capital growth. The number of stock exchanges and companies listed has grown over the years making stocks one of the most common investment choices for both individual and institutional investors. One of the biggest challenges for investors is identifying the right stocks to buy in order to achieve their financial goals. Stock prices are often a reflection of the sentiments investors have about the underling business. Positive sentiments often drive the stock price in a positive direction while negative sentiments result in drops in the stock.

There are many factors that influence how the market views a company. Fundamental analysis looks at different measures within a company to provide a picture of its health. Companies trading on public exchanges are required to make disclosures on their business activities. These include records of the company’s performance in a reporting period (often a year) and the overall position of the company as at a given time. The reports provide investors with a view of the company’s cash flows, investments, financing, ownership structure, management among others.

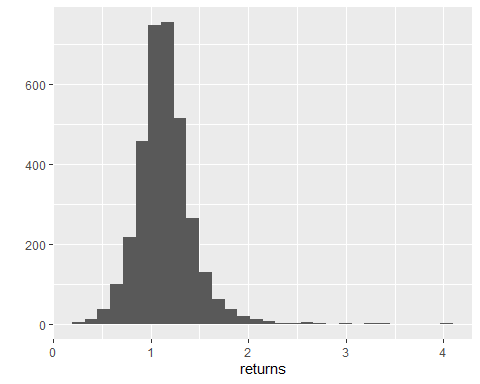
The goal of this analysis is to evaluate the relationship between the different records of activities within a company and the changes in stock prices over a given period of time. The analysis evaluates the hypothesis that changes in the level of business activities such as sales, profitability, financing and investment from one period to another have a linear relationship with the change in the company’s stock price during the same period. The assumption therefore, is that increasing levels of assets and business activities such as sales and investment would reflect positively on the company leading to increased stock prices while increasing liabilities in a given period would lead to a reduction in stock price. The analysis also checks whether there is a difference between the change in stock prices for companies in different sectors. This is because there are many other factors external to a business that influence its operations. These may include the regulatory environment, competition, extreme natural and economic events among others. Such factors often have a similar effect on companies operating within the same industry. A significant difference in the change in stock prices between different sectors would give way for us to evaluate whether a linear relationship is significant by including it as one of the predictors in the regression model for change in stock prices.

# Data description

Data on the fundamentals of 448 stocks and daily closing prices for 501 stocks are available in different datasets. The data are imported as *fundamentals* and *price\_adjusted* respectively. A record on the sector in which 505 companies operate is also available in a different dataset imported as *securities*. The fundamentals and securities datasets are merged appropriately based on the stocks for which both the records are available.

NOTE: Before proceeding with the analysis, an additional observation is added to the data. The dataset currently has 1781 observations with 86 features. The new observation has the values for each of the continuous features to the mean of the respective features. For example, Common.Stocks is the average common stock for the 1781 observations and the same is done for all the continuous data. The observation is assigned the GICS.Sector, "Consumer Staples" which is the most dominant sector in the dataset. The GICS sub sector is chosen randomly as "Agricultural Products" from the consumer staples sector. The new observation is assigned an arbitrarily chosen value in remaining columns with categorical variables. The value is chosen from the records already present in the respective columns. The new observations are assigned to a stock name "New observation" with the symbol "NEW". The result is now a dataset with 1782 observations. The new observation "NEW" is also added to the stock prices data as a single point with prices being the average price in the respective columns.

From the prices dataset, the adjusted daily closing prices are used to obtain the annual change in stock price. This is the price adjusted for stock splits. The annual return is obtained from the following formula:



The annual returns for the 501 stocks provided in the dataset are normally distributed with a mean and standard deviation of and respectively. There are 75 features describing the fundamentals of 449 stocks, each for a couple of years, and add a new observation resulting in a total of 1782 observations. For each of the stocks, the change in the level of the features is obtained by subtracting the records presented for consecutive years. The annual change for each feature is obtained as a compounding factor for the feature using the formula which can be expressed fully as:

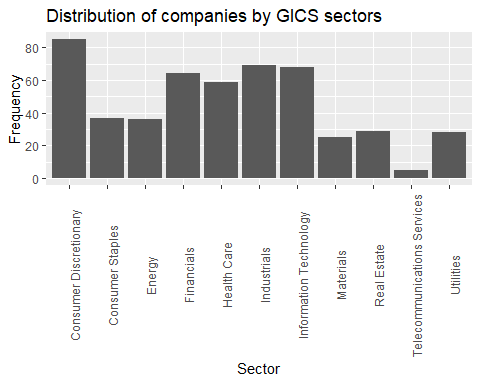
where represents the year for which the change is calculated and represents the stock being evaluated. Annual returns and the transformed features are then merged into a single dataframe for evaluation of missing values. First, features that had more than of their records missing were removed completely from the dataset. With the remaining dataset, observations with a missing record for any feature were removed. The result is a clean dataframe with 43 fundamentals describing 274 companies in 756 observations.

# Methods

Correlation is a measure of the extent to which two variables are related. This is important in regression analysis since it influences the decision on which features are best used for prediction. The amount of variation in a response that is explained by the linear combination of predictive features is the R-squared value, whose square root gives the correlation coefficient between the response and the predictors. Features therefore that have a high correlation with the response are desired when fitting a regression line to the data. It is however possible to have features that are significantly correlated. This suggests that their relationship is significant and therefore cause confounding when both of them are used in the same model. Confounding is where the apparent relationship between the response and a feature is actually influenced by another feature in the model. Confounding is prevented by identifying features that have a high correlation, and removing them from the list of predictors.

The fundamentals *Gross.Profit* is flagged to be removed due to its high correlation with the *Pre.Tax.ROE*. *Total.Liabilities…Equity* is highly correlated with both *Total.Liabilities* and *Total.Assets* and it is therefore removed. The same treatment is given to *Total.Assets* leaving only *Total.Liabilities* to be used in the model. *Total.Revenue* is correlated with the *Cost.of.Revenue* while *Pre.Tax.Margin* is correlated with the *Gross.Margin* and *Pre.Tax.ROE*. The *Pre.Tax.ROE* is also correlated with *Gross.Profit* and it is also removed. Finally, net income is removed for its high correlation with *Net.Income.Applicable.to.Common.Shareholders*.

The industry in which a company operates is also evaluated to see if it results in a difference in the annual change in stock price. Below is the distribution of companies by sector for the entire set of companies available in the dataset. The consumer discretionary sector has the highest number of companies in the stock market while telecommunications has the smallest representation. Industries, information technology, financials and healthcare also have a significantly higher number of companies trading their stocks in the exchange compared to the other sectors. In general, however, we may conclude that there is adequate representation of companies in each sector for the analysis.



# Results

First the annual change in stock price is evaluated for differences between the sectors.

|  |  |  |
| --- | --- | --- |
| **GICS.Sector** | **mean** | **SD** |
| **Financials** | *1.218999* | *0.206644* |
| **Health Care** | *1.192604* | *0.250041* |
| **Information Technology** | *1.167652* | *0.333930* |
| **Consumer Staples** | *1.133040* | *0.189811* |
| **Industrials** | *1.129681* | *0.271772* |
| **Utilities** | *1.111801* | *0.180572* |
| **Real Estate** | *1.107736* | *0.213395* |
| **Consumer Discretionary** | *1.057922* | *0.242777* |
| **Materials** | *1.047804* | *0.220716* |
| **Energy** | *1.014876* | *0.322346* |
| **Telecommunications Services** | *1.009426* | *0.211374* |

Companies in the financial industry had the highest average annual change in stock price growing by an average of . This is followed by the Healthcare and Information technologies companies respectively. Companies operating within the telecommunications, energy and materials sectors had the lowest annual change in stock price with growth of less than each. Information technology and energy stocks were the riskiest while consumer staples and utilities were the least risky. The Kruskal-Wallis test for independence is used to evaluate whether the difference in average annual change in stock prices between the industries is significant.

The test returns a p-value of 0.000381607810 which is . This tells us that their is a significant difference in returns between the different sector. The sectors whose average annual change in stock prices are not significantly different are represented by the same letter.

|  |  |  |
| --- | --- | --- |
| **Group** | **Letter** | **MonoLetter** |
| **ConsumerDiscretionary** | a | a |
| **ConsumerStaples** | ab | ab |
| **Energy** | a | a |
| **Financials** | ab | ab |
| **HealthCare** | b | b |
| **Industrials** | ab | ab |
| **InformationTechnology** | ab | ab |
| **Materials** | ab | ab |
| **RealEstate** | ab | ab |
| **TelecommunicationsServices** | ab | ab |
| **Utilities** | ab | ab |

Annual changes in stock prices in the consumer discretionary and energy sectors are not significantly different. The healthcare sector has annual stock price changes that are significantly different from all the other sectors. The remaining sectors all have annual stock price changes that are not significantly different. Due to the significance in the differences, the sector a company operates in is included in the regression analysis.

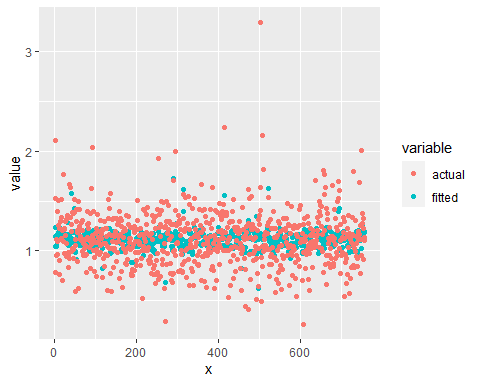
To select the best features from the other 43 features, a regression line is fit with the % annual change in stock price as the response and fundamentals whose parameters are found to be insignificant i.e. are removed. The following table shows the fundamentals whose parameters were found to be significant. These are the features that are selected and fitted to obtain the regression equation to estimate the annual change in stock prices.

|  |  |
| --- | --- |
|  | Pr(>|t|) |
| (Intercept) | 1.59E-12 |
| Common.Stocks | 6.71E-03 |
| Net.Cash.Flow.Operating | 1.01E-02 |
| Operating.Margin | 2.54E-03 |
| Other.Equity | 5.12E-03 |
| Other.Operating.Activities | 3.51E-02 |
| Retained.Earnings | 4.73E-02 |
| GICS.SectorConsumer Staples | 4.61E-02 |
| GICS.SectorFinancials | 1.64E-02 |
| GICS.SectorHealth Care | 1.53E-04 |
| GICS.SectorIndustrials | 3.45E-02 |
| GICS.SectorInformation.Technology | 1.42E-03 |

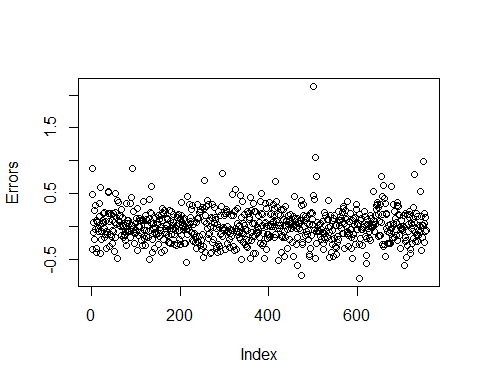
We note that the model identifies 7 features, including the sector, as significant predictors of the change in stock prices. Since industry is a categorical predictor, a parameter is estimated for each of its levels resulting in the additional estimates as seen in the table above. The regression model is refitted with the significant fundamentals only. The parameter estimates and their corresponding statistics are shown.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Estimate** | **Std. Error** | **t value** | **Pr(>|t|)** |
| **(Intercept)** | *1.004511* | *0.02634* | *38.13622* | *7.7E-177* |
| **Common.Stocks** | *-6.26E-05* | *2.36E-05* | *-2.65103* | *0.008196* |
| **Net.Cash.Flow.Operating** | *0.026126* | *0.008702* | *3.002347* | *0.002769* |
| **Operating.Margin** | *0.015825* | *0.005106* | *3.099369* | *0.002013* |
| **Other.Equity** | *0.001781* | *0.000609* | *2.922179* | *0.003582* |
| **Other.Operating.Activities** | *0.001363* | *0.000579* | *2.354597* | *0.018803* |
| **Retained.Earnings** | *0.003556* | *0.001544* | *2.303333* | *0.021537* |
| **GICS.SectorConsumer Staples** | *0.079417* | *0.039067* | *2.032853* | *0.042424* |
| **GICS.SectorEnergy** | *-0.05925* | *0.039827* | *-1.48761* | *0.137279* |
| **GICS.SectorFinancials** | *0.16036* | *0.069936* | *2.292958* | *0.02213* |
| **GICS.SectorHealth Care** | *0.132701* | *0.034788* | *3.814519* | *0.000148* |
| **GICS.SectorIndustrials** | *0.063679* | *0.031539* | *2.01904* | *0.043843* |
| **GICS.SectorInformation Technology** | *0.111806* | *0.03386* | *3.302001* | *0.001006* |
| **GICS.SectorMaterials** | *0.008545* | *0.041238* | *0.207212* | *0.835901* |
| **GICS.SectorReal Estate** | *0.049438* | *0.054365* | *0.909379* | *0.363446* |
| **GICS.SectorTelecommunications Services** | *-0.05019* | *0.084085* | *-0.5969* | *0.550756* |
| **GICS.SectorUtilities** | *0.058966* | *0.044973* | *1.311138* | *0.190218* |

The regression model fitted is a good representation of the relationship between the annual change in the selected fundamentals: common stock, net operating cashflow, operating margin, other equity, other operating activities, retained earnings and the industry: and the change in stock prices for the same period. An R-square value of however shows low predictive power for the regression model since the fitted fundamentals only explain of the variation in annual changes in stock prices.



The scatter plot above gives shows the fitted values from the regression model alongside the actual values. The plot of the residuals shows no evidence of heteroscedasticity as they are well distributed along the 0 line for all the observations. This is more evidence to support that the model captures the relationship between the response and the predictors well.

  
**Conclusion**

The analysis set out to investigate the significance of a linear relationship between predictors derived from the companies’ fundamentals with the change in stock price for the given company over the same period of time. The analysis finds that the annual changes in common stock, net operating cashflow, operating margin, other equity, other operating activities, retained earnings and the industry have a significant relationship with the changes in stock price and they are used to estimate the model parameters. The model is good at the identification of significant predictors but has a low predictive power suggesting that many other factors exist that may be used to explain the remaining variation in the stock price changes. The errors satisfy the normality assumption which is an important factor in determining whether a model is a good fit for the set of data.

The industry within which a company operates is identified to be significant in determining how prices change. This confirms the assumption set out earlier in the report that external factors that influence the change in prices tend to have a similar effect on companies within the same sector since these companies are likely to react the same when exposed to the same force.