Scientific
Research
Publishing

Analytical Predictive Modeling: Impact of Financial and Economic Indicators on Stock

Jayanta K. Pokharel*, Erasmus Tetteh-Bator, Chris P. Tsokos

Department of Mathematics and Statistics, University of South Florida, Tampa, FL, USA
Email: *jpokharel@usf.edu, etettehbator@usf.edu, ctsokos@usf.edu

Abstract

Stock price prediction is considered as an important task and is of great attention as predicting stock prices successfully may lead to attractive profits for investors. Information Technology Sector of S&P 500 is one of the most sought after business segments in S&P 500 and is one of the most attracting areas for many investors due to high percentage annual returns on investment over the years. We used Microsoft Corp. (MSFT), one of the leading companies of the Information Technology Sector Index of S&P 500 information to build a non-linear real data-driven analytical model which accurately predicts the Weekly Closing Price (WCP) of the stock with predictive accuracy of 99.3% using six financial, four economic indicators and their two way interactions as the attributable entities that drive the stock returns. We rank the statistically significant indicators and their interactions based on the percentage of contribution to the WCP of the stock that provides significant information for the beneficiary of the proposed predictive model. We present a unique way for feature selection when multicollinearity presents in the multiple regression dataset using L1-regularization based on supervised machine learning algorithm.

Keywords

Financial and Economic Indicators, Financial Instruments, Johnson Transformation, Multicollinearity, L1-Regularization, Homoscedasticity, k-Fold Cross Validation

1. Introduction

Investors are highly attracted to the stock market due to unlimited profit return possibilities. However, high returns are possible at a cost of high risk, thus, a stock market is called risk-return trade-off [1]. The fluctuations of stock prices

affect the investor's decision on investing their capital. In the financial markets, stock price movements are governed by the stock price index and it is considered as the references for investors to invest in the capital markets.

The S&P 500 index is widely considered to be one of the best single gauges for the U.S. equity markets. It does not only reflect the economic activities of the United States but also has a greater impact on global economy. S&P 500 consists of 11 business sectors and each of them has their own price index. The Information Technology Sector is one of the most sought after business sectors of S&P 500 due to the presence of tech-giants like Apple Inc., Microsoft Corp., Nvidia Corp., etc. It is estimated that in 2021, the United States tech-sector contributed around 1.8 trillion dollars to the country's GDP, making up approximately 9.3% of total GDP [2]. The best performing Sector in the S&P 500 in the last 10 years is the Information Technology Sector that recorded 16.17% annualized return followed by the Healthcare, 13.9% [3].

Stock price prediction is considered as an important task and is of great attention as predicting stock prices successfully may lead to attractive profits for investors. In this context, we proposed a high quality real-data driven predictive model based on one of the leading companies of the Information Technology Sector Index of S&P 500, that is, Microsoft Corp. (MSFT) and the detailed of the stock selection process are explained in Subsection 2.1. The Weekly Closing Price (**WCP**) of the MSFT stock is used as a measure of our "**response**". In our model building process, we have considered six financial indicators (Beta, Free Cash Flow Per Share, Price-to-Book Ratio, Price Earning Ratio, PEG Ratio & Dividend Yield), four economic indicators (Interest Rate, Index of Consumer Sentiment, US Personal Saving Rate & US GDP) and their two way interactions that can possibly contribute to the WCP of the MSFT stock. The detailed explanation about our attributable indicators is presented in Subsection 2.2.

The attributable entities (indicators) that are included in the proposed predictive model have a significant relevance in the literature of finance and economics. Tang and Shum in their study on "Conditional Relations between Beta and Return", found that a measure of stocks volatility of returns which is defined as Beta-factor to have significant effect on returns [4]. However, a study conducted by Sherelene Enriquez-Savery shows that the Beta-risk factor is wrongly calculated in practice [5]. Many researches and business analysts believe that Free Cash Flow Per Share (FCF/Share) and Dividend Yield (Div Yield) are closely associated and play a crucial role in the stock returns. Also, stocks with high dividend yields usually have an advantage of being attractive to many investors due to regular return advantage over their lower-yielding counterparts. However, our study shows that FCF/Share does not statistically significantly contribute but Dividend Yield does contribute to the WCP of the MSFT stock. People believe that Price-to-Book (P/B) Ratio indicates whether the stock is overvalued and can be used in determining the best value of the stock at a given time. Our research shows that interaction of the P/B Ratio with other indicators significantly contribute to the Weekly Closing Price of the stock. Many researchers believe that

Price-to-Earning (P/E) Ratio and Price-Earnings-to-Growth (PEG) Ratio are two of the most crucial indicators that influence the stock returns. Lagevardi [6] research supports that P/E ratio is more directly related to stock returns than the PEG Ratio and thus, stock return of the companies is more affected by the P/E ratio as compared to PEG Ratio. On the other hand, many researchers argue that PEG Ratio gives more complete picture of the stock returns as it also accounts for expected future growth. Our study shows that PEG Ratio and its interactions with other indicators are statistically powerful contributors in determining the WCP of the stock price.

Prior studies have shown that changes of interest rate have a significant negative relationship with changes of the stock price [7]. We found that interest rate alone is a weak attributable entity in determining the WCP of the stock, however, its interactions with other indicators has statistically significant effect on the stock price. Similarly, many researchers and financial analysts believe that good financial market environment motivates the investors' interest to keep the growth momentum of the stock market. A study by Lemmon *et al.* (2006) has shown that investors' confidence level exhibits forecasting power for the stock returns [8]. Our findings from this study also support this argument. Other economic indicators which are used in the proposed model are, the US Personal Saving Rate (PSR) and the US GDP. Studies have shown that current saving rates influences future consumption and supports investments [9], thus, it is important to understand its impact on stocks/index price. Our study also supports the fact that PSR is a statistically significant indicator in predicting stock price. A study of interlinakge between stock market and GDP growth shows that there is no direct connection between stock market growth and the GDP growth [10]. However, in our study we found that the US GDP is the number one attributable entity in determining the WCP of the MSFT stock.

In this paper, we propose a highly accurate real data-driven predictive model with predictive accuracy of 99.3% to predict the WCP of MSFT stock. We also conducted a comparative study of the significant effect of financial and economic indications, and their interactions between MSFT stock and the Information Technology Sector Index of S&P 500 [11] on their respective weekly closing price. The article presents some intriguing findings and, the proposed model's usefulness. The detailed methodology and procedure of the proposed model building process are discussed in the following section.

2. Methodology

The detailed procedure and the methodology of our proposed predictive model building process are presented in Subsection 2.1, 2.2 and 2.3.

2.1. Selection of Appropriate Stock

In this study, our main goal is to perform our analysis for a particular stock based on the Information Technology Sector Index of S&P 500 and we used Microsoft Corp. (MSFT) stock to build our predictive model. Microsoft Corp. is

ranked number 2 among 75 companies listed in the Information Technology Sector Index of S&P 500, which is only behind Apple Inc. (AAPL) in the index based on their market capitalization and revenue. During our study period 2017-2019, Microsoft Corp. outperformed Apple Inc. in key areas of interest for investors such revenue growth, net profit growth, dividend yield etc. [12]. Microsoft Corp. has been one of the most reliable companies with consistent growth and excellent financial performance in its long history with average volatility among the tech-sector's stocks, and it is being one of the most attractive stocks for investment for many reasons, we chose MSFT stock information from the Information Technology Sector Index of S&P 500 to build our proposed predictive model.

2.2. Data and Description of the Indicators

After we select MSFT as the leading company, we collected the required information about our selected indicators using various sources such as Yahoo Finance, FRED Economic Data, Zacks Investment Research, Alpha Query and Morningstar.com. The data set includes information from January 2017 to December 2019. The weekly data information is used to structure the required data base. After extensive literature review of the subject area, six financial indicators and four economic indicators are considered that may influence the Weekly Closing Price (**WCP**) of the stock as a measure of our response.

In addition, for the convenience of the readers, we define below the financial indicators from number 1 - 6, and economic indicators from number 7 - 10 that are significant entities of our analytical model.

1) **Beta** (X_1): The beta value is a statistical measure that compares the volatility of return of a specific stock in relation to those stocks of the market as a whole. In general, stocks with higher beta value are considered to be riskier, thus, investors will expect higher returns. That is,

$$\text{Beta} = \frac{Cov(R_i, R_m)}{Var(R_m)},$$

where R_i is the return on individual stock and R_m is the return on overall market. $Cov(.,.)$ is the covariance between R_i and R_m that measures the changes in stock's returns with respect to the changes in market's returns, $Var(.)$ is the variance of the excess market returns over the risk-free rate of returns.

2) **FCP/Share** (X_2): Free Cash Flow Per Share (FCF/Share) is a measure of a company's financial flexibility [13]. It calculated by dividing Free Cash Flow of the company by the total number of shares outstanding. That is,

$$\text{FCF/Share} = \frac{\text{Free Cash Flow}}{\text{Number of Share Outstanding}}.$$

3) **P/B Ratio** (X_3): Price-to-Book (P/B) Ratio compares a company's current market value to its book value [14]. And the book value is defined as the value of

all assets minus liabilities owned by a company. That is,

$$\text{P/B Ratio} = \frac{\text{Market Value Per Share}}{\text{Book Value Per Share}}.$$

4) **P/E Ratio** (X_4) Price-to-Earnings (P/E) Ratio is the ratio that measures the current price of a stock concerning its earnings per share [15] and is given by

$$\text{P/E Ratio} = \frac{\text{Current Price Per Share}}{\text{Earnings Per Share}}.$$

5) **PEG Ratio** (X_5) Price-Earnings-to-Growth (PEG) Ratio is the stock's Price-to-Earnings (P/E) Ratio divided by the growth rate of its earnings for a specified period [16] and is given by

$$\text{PEG Ratio} = \frac{\text{P/E Ratio}}{\text{Annual EPS Growth}}.$$

6) **Dividend Yield** (X_6): Dividend Yield (Div_Yield) is the percentage measure of the company's share price that it pays out in dividends each year and is given by

$$\text{Dividend Yield} = \frac{\text{Annual Dividend Per Share}}{\text{Current Share Price}}.$$

7) **Interest Rate** (X_7): The US Federal Fund Rate is used. It is the target interest rate set by the Federal Open Market Committee [17]. This target is the rate at which the Fed suggests commercial banks borrow and lend their excess reserves to each other overnight.

8) **US ICS** (X_8): The Michigan Survey Research Center has developed the Index of Consumer Sentiment (ICS) to measure the confidence or optimisim (passimism) of consumers in their future well-being and coming economic condition [18]. The ICS measures short and long-term expectations of business conditions and the individual's perceived economic well-being.

9) **US PSR** (X_9): The US Bureau of Economic Analysis (BEA) publishes the US Personal Saving Rate [19]. The US Personal Saving Rate is the personal saving rate as a percentage of personal income. It is percentage measure of individuals' income left after they pay taxes and expenditures.

10) **US GDP** (X_{10}): Gross Domestic Product (GDP) of the United States (in billion) is used. GDP is defined as the measure of monetary value of all finished goods and services made within a country during a specific period [20]. The components of GDP include personal consumption expenditures (C), business investments (I), government spending (G), exports (X), and imports (M). That is,

$$\text{GDP} = C + I + G + (X - M).$$

2.3. Development of Statistical Model

To develop our multiple regression predictive model, there are several statistical assumptions that our data must satisfy. Then, we proceed to verify these as-

sumptions.

Normality of the Weekly Closing Price (WCP): In developing the proposed statistical model for the MSFT stock Price as a function of the attributable indicators, one of the main assumptions is that the response indicator "**WCP**" should follow the Gaussian Probability Distribution. Normal QQ plot and the normality tests are used to verify the normality of our response, **WCP**.
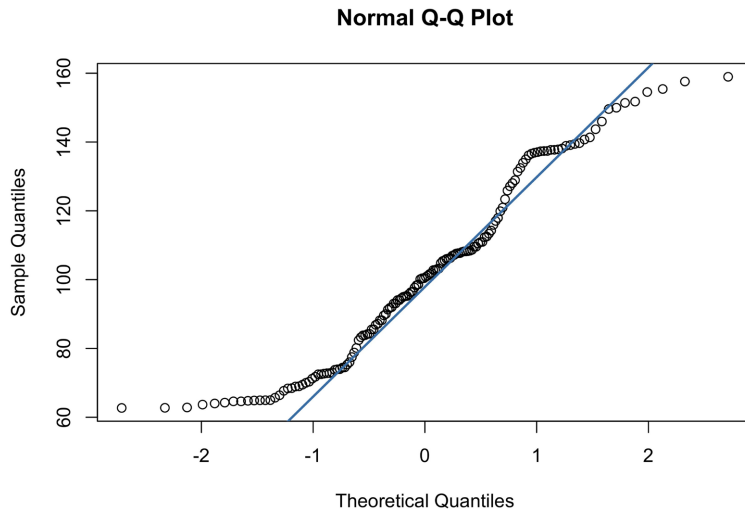
Normal Q-Q Plot

Figure 1. Normal Q-Q plot of WCP.

Figure 1 above, shows that there is systematic deviation from normality and the WCP does not entirely follow a Gaussian Probability Distribution. The goodnesss-of-fit test using Shapiro-Wilk normality test yields a p-value = $2.4e^{-05}$ which is less than 0.05 at 5% level of significance. Also, the Anderson-Darling normality test has the p-value = $3.846e^{-05}$. Both of these normality tests confirm that the response variable **WCP** does not follow the Normal Probability Distribution. Therefore, the Normal Q-Q Plot of the **WCP** supports the fact that natural phenomena such as the Weekly Closing Price of the stock do not follow the Gaussian Probability Distribution.

In the process of developing a statistical model, our main goal is to express our response WCP in terms of non-linear mathematical function of all significantly contributing indicators including their interactions with high degree of accuracy. Thus, one of the pure forms of the model with all possible interactions and additive error terms that can possibly estimates the WCP of the stock is expressed as follows:

$$\text{WCP} = \beta_0 + \sum_j \beta_j X_j + \sum_{j \neq l} \gamma_{jl} X_j X_l + \varepsilon \tag{1}$$

where,

- β_0 is the intercept of the regression model,
- β_j is the coefficient of jth individual indicator X_j,
- γ_{jl} is the coefficient of jlth interaction term $X_j X_l$,

- $\varepsilon \overset{\text{iid}}{\sim} N\left(0, \sigma^2\right)$ is a Gaussian error terms (residual error).

The main assumption behind constructing the above model is that the response indicator "**WCP**" should follow the Gaussian Probability Distribution. However, we noticed that WCP does not support it. Therefore, we must apply a non-linear transformation to WCP and see if the transformation can adjust the scale of the response to follow the Normal Probability Distribution. We used the Johnson Transformation bounded family [21] to transform WCP. Let X be a continuous random variable whose distribution is unknown and is to be approximated, three normalizing transformations as proposed by Johnson has the following general form:

$$Z = \gamma + \delta \ln\left[\frac{X - \xi}{\lambda + \xi - X}\right], \xi < X < \xi + \lambda \tag{2}$$

where,

$-\infty < \gamma < \infty$, γ —the shape parameter,

$\delta > 0$, δ —the shape parameter,

$\lambda > 0$, λ —the scale parameter,

$-\infty < \xi < \infty$, ξ —the location parameter.

After we applied Johnson Transformation (2) on WCP, we obtained Equation (3) to estimate the new Transformed Response Indicator (WCP$^{\text{T}}$).

$$\text{WCP}^{\text{T}} = 0.4293 + 0.6113 \ln\left[\frac{X - 61.3432}{161.2604 - X}\right], \tag{3}$$

where $\gamma = 0.4293$, $\delta = 0.6113$, $\xi = 61.3432$ and $\lambda = 99.9172$.

We then checked the normality of the WCP$^{\text{T}}$ for goodness-of-fit using Shapiro-Wik test. The normality test confirms that the new transformed response indicator WCP$^{\text{T}}$ does follow the Normal Probability Distribution.

Here onward, we use the transformed response indicator "WCP$^{\text{T}}$" as the new response indicator to conduct our statistical analysis and later, we must apply anti-Johnson transformation to obtain the original scale of WCP.

Non-Linear Relationship among Indicators: Multiple linear regression assumes that there is little or no multicollinearity in the data and correlation analysis is an important part of the model building process. Figure 2 below, shows the strength of linear association among all possible attributable indicators.

There should be non-linear relationship among our attributable indicators to run the multiple regression model. We observed that most of our attributable indicators are linearly associated with WCP of the stock which is a good sign to develop a multiple regression model for the given dataset. However, we also inspected that there is strong linear association among some indicators showing the possibility of multicollinearity in our dataset. Given the cut off value of correlation coefficient between 0.80 to −0.80, GDP is found to be strongly correlated with Beta, P/B Ratio, P/E Ratio, Dividend Yield and Interest Rate. Also, P/B Ratio is strongly correlated with Beta and P/E Ratio. In addition Dividend Yield is highly correlated with Beta, P/B Ratio and P/E Ratio.
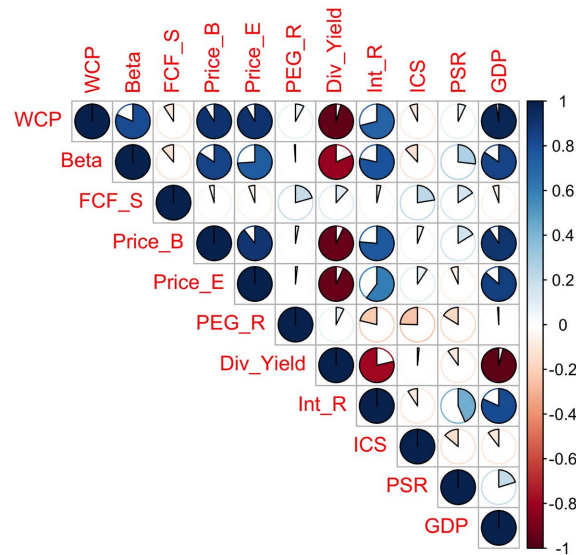
Figure 2. Correlation matrix of all possible attributable indicators.

We then checked the Variance Inflation Factor (VIF) by constructing a regression model to statistically verify the existence of multicollinearity among our attributable indicators. VIF is a measure of the amount of multicollinearity present in a set of multiple regression variables. The VIF score less than 10 indicates that the dataset has no multicollinearity effect among the attributable indicators, otherwise they are linearly correlated.

We observed from **Table 1** below, that VIF value for P/B Ratio (X_1), P/E Ratio (X_4), Dividend Yield (X_6) and GDP (X_6) are 12.12, 16.68, 46.12 and 20.53, respectively. These VIF values are the strong evidence of the presence of multicollinearity in our dataset. We dropped these indicators with VIF values greater than 10 one at a time and checked the multicollinearity effect. Even by doing so, we have not observed significant change on correlation status among the remaining indicators. We believe that the above mentioned indicators with high VIF values have a significant importance in our model building process, it would not be a good idea to drop them from our initial model. In this condition, we can not use multiple regression to build our model.

Table 1. Variance Inflation Factor (VIF).

X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}
5.13	1.62	12.12	16.68	1.71	46.12	7.12	1.61	1.78	20.53

We then proceed to treat the multicollinearity present in the dataset through feature selection using lasso regression and details are presented below.

Lasso Regression for Feature Selection: LASSO stands for Least Absolute Shrinkable and Selection Operator and can also be called L1-regularization or L1-norm. When we compare it with linear regression, Lasso is different in the sense that it uses penalty term in its equation to penalize the highly correlated

covariates and selects a reduced set of useful covariates for a model [22]. Lasso regression works based on supervised machine learning algorithm that shrinks the coefficients of determination towards zero to prevent from overfitting in the model [23]. We know that linear regression gives the regression coefficients as observed in the dataset where as the lasso regression regularizes these coefficients to avoid overfitting, thus it works better on different datasets. Lasso regression facilitates variable selection when multicollinearity is present in the dataset by penalizing less important variables and enhances feature selection for a simple model creation [24].

The fundamental notion underlying the LASSO is to optimise the tradeoff between bias and variance in regression estimation where bias refers to the deviation between predicted and actual values and variance to the variability in prediction. The LASSO thus optimises the trade off between accuracy and consistency. Traditional ordinary least square (OLS) estimates minimise the residual sum of squares thereby reducing bias but at the cost of higher variance. Thus, for LASSO, a small modification is required to the cost function of the OLS as shown below.

$$J(m) = \alpha + \sum_{i=1}^{N} \left(y_i - \sum_{j} x_{ij}\beta_j \right)^2 + \lambda\sum_{j=1}^{p} |\beta_j|, \tag{4}$$

where, x_{ij} is the standardized predictors and y_i is the centered response value. The term $\lambda\sum_{j=1}^{p} |\beta_j|$ in the Equation (4) is called "Penalty Term". In a ideal situation, in order to have sum of square errors zero, the best-fit line must pass through all data points. The additional term $\lambda|slope|$ should be minimized to minimize the cost function. Minimum slope produces less steeper lines such that all data points does not fall on the line and this will help to prevent from over-fitting the model. λ is a positive real number and we call it a regularization parameter in the model. If λ is too high, this will minimize or "shrink" the slope to zero. That is, it suppresses the coefficients of highly correlated features and makes the model simple. The regularization parameter (λ) can be determined by cross-validation method and it avoids under-fitting (when λ is too high) and over-fitting, (when λ is too small).

To determine the optimal value of the regularization parameter λ that minimizes Mean Squared Error, we applied k-fold cross validation method for the train dataset. The sequence of lambda vector (500) are created and repeated 3 times over $k = 5$, that is 5 - fold cross validation using cv.glmnet package in R. The optimal value for λ is found to be $9.593209e^{-05}$, that is, $\log(\lambda) = -9.25187$). We employed lasso regression model with 10 indicators and their all possible two way interaction terms ($\binom{n}{x} = 45$, where $n = 10$ and $k = 2$).

Figure 3 below, illustrates the k-fold cross validation for optimizing regularization parameter λ. We observeed that MSE increases with increasing value of λ and we acheive the optimal λ value, that is λ minimum that minimizes the MSE when $\lambda = 9.160598e{-05}$. Similarly, **Figure 4** below, illustrates how the coefficient shrinkage happens in L1 regularization with increasing value of λ. We ob-

served that for optimal value of lambda, that is, $\log(\lambda) = -9.25187$, lasso model retains 9 attributable indicators and 24 interaction terms and shrinks the remaining coefficients of the individual indicator and interactions to zero to exclude from the model.
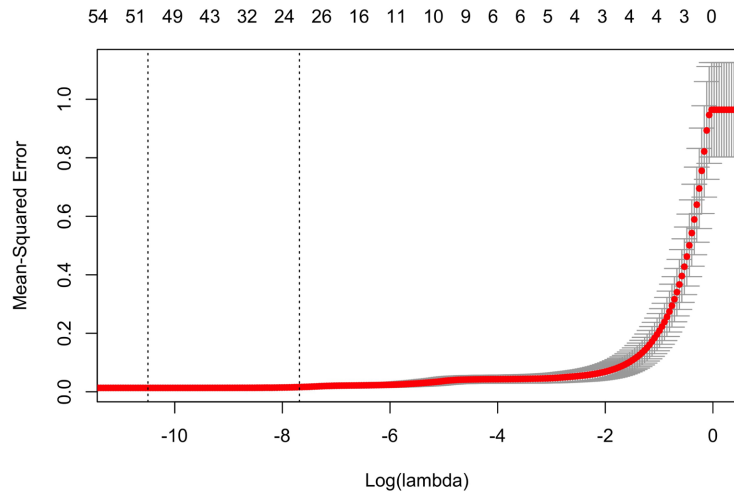

Figure 3. Graph of optimized λ value using K-Fold cross validation.
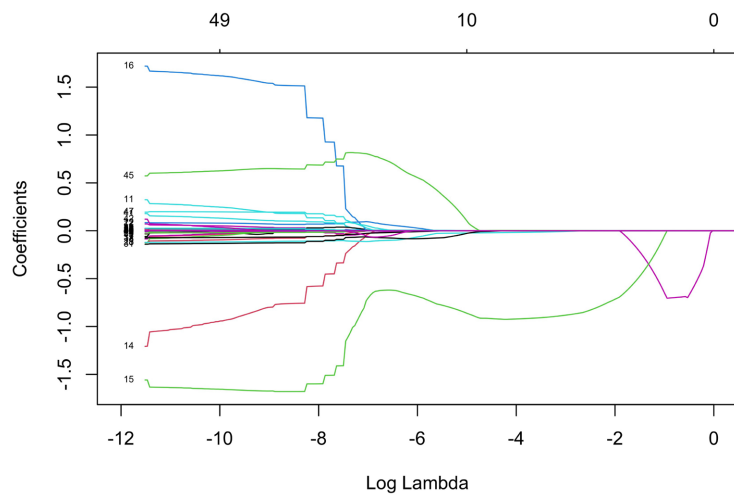

Figure 4. Graph to show "coefficient shrinkage" via L1 regularization.

We then proceeded to fit the multiple linear regression model using the attributable indicators and their interactions selected by lasso regression model.

2.4. Fitting the Statistical Model

During our data preprocessing step, we observe that our attributable indicators are in different scale and range. Therefore, feature scaling (the process of normalizing the range of features/column vectors of different units in a dataset) is used to standardize the range of our attributable indicators. We randomly split the data into train and test dataset in the ratio of 80:20. As a statistical model building process on the train dataset, we proceeded to estimate the coefficients

(weights) of the actual contributable indicators for the transformed data in Equation (3). We run multiple regression model including all individual indicators and their two way interaction terms selected by lasso regression model. It is important to note that we used all 10 individual indicators in the model because lasso regression discovered their interactions as important entities for the model building process. Backward elimination is deemed one of the best traditional methods for a small set of feature vectors to handle the problem of overfitting [25], we then proceeded to determine the most significant individual indicators and their interactions using stepwise backward elimination method. After careful considerations, we noticed that five attributable indicators and eleven interaction terms remained statistically significant in our final model. In our final multiple regression model, we observed that Beta, P/E Ratio, PEG Ratio, PSR and GDP and, the interaction terms Beta $\cap$ P/B Ratio, Beta $\cap$ PEG Ratio, P/B Ratio $\cap$ P/E Ratio, P/B Ratio $\cap$ Div _ Yield, P/E Ratio $\cap$ Int_ Rate, PEG Ratio $\cap$ Div _ Yield, Div _ Yield $\cap$ Int_ Rate, ICS $\cap$ PSR and ICS $\cap$ GDP are found to be highly statistically significant where as the interaction of P/B Ratio $\cap$ Int_ Rate and Int_ Rate $\cap$ ICS are moderately significant. The FCF/Share and its interactions with other indicators are found to be non-significant at 5% level of significance, thus excluded from the model. Similarly, P/B Ratio, Dividend Yield, Interest Rate and ICS are not individually significant where as there interaction with other indicators are found to statistically significant at 5% level of significance. The trained model has the R-squared of 0.993 and the adj-R-squared of 0.992. The high R-Squared value indicates that the proposed model is of high quality model with excellent predictive accuracy of 99.3%.

The best preferred statistical model with all significantly contributable indicators and their interactions that estimates the Transformed Weekly Closing Price ($\widehat{\text{WCP}^{\text{T}}}$) of the MSFT stock is given by Equation (5).

$$
\begin{aligned}
\widehat{\text{WCP}^{\text{T}}} = {} & 0.0931 - 0.1304X_1 + 0.1557X_4 + 0.0501X_5 - 0.0498X_9 \\
& + 0.7118X_{10} + 0.2064X_1X_3 - 0.0933X_2X_5 + 0.2756X_3X_4 \\
& + 0.3831X_3X_6 - 0.1408X_3X_7 - 0.2938X_4X_7 - 0.1186X_5X_6 \\
& - 0.2744X_6X_7 - 0.1223X_7X_8 + 0.0598X_8X_9 + 0.1276X_8X_{10}.
\end{aligned}
\tag{5}
$$

Equation (5) estimates the Transformed Weekly Closing Price (WCP$^{\text{T}}$) of the stock based on Johnson transformation. Here, anti-transformation on Equation (3) is required to estimate the desired prediction of WCP of the stock, and it is given by the following Equation (6) or (7).

$$
\widehat{\text{WCP}} = \xi + \frac{\lambda}{1 + \exp\left(\dfrac{\gamma - \widehat{\text{WCP}^{\text{T}}}}{\delta}\right)}
\tag{6}
$$

Or,

$$
\widehat{\text{WCP}} = 61.3432 + \frac{99.9172}{1 + \exp\left(\dfrac{0.4293 - \widehat{\text{WCP}^{\text{T}}}}{0.6113}\right)}.
\tag{7}
$$

Equation (7) represents the best preferred statistical predictive model to estimate the Weekly Closing Price (WCP) of the MSFT stock. Given any standardized values of the attributable indicators, the Equation (5) estimates the $\widehat{\text{WCP}^{\text{T}}}$ and we substitute the value of $\widehat{\text{WCP}^{\text{T}}}$ in the analytical model represented by Equation (7) to predict the Weekly Closing Price of the stock.

3. Evaluation of the Predictive Model

In this section we proceed to evaluate the quality of the proposed model based on model performance and assumptions.

Model Performance: The proposed analytical model is evaluated based on various performance metrics and results are tabulated in Table 2, below.

The R^2 of the proposed model is 99.3% which is pretty closed to adj-R^2, 99.2%. The closer the R-squared and adjusted R-squared value are, the more accurate the model is, that insures the predictive accuracy of the proposed model. It has root mean squared error (RMSE) of 2.91 and mean absolute percentage error (MAPE) of 1.99%. We also used relative root mean square error (RRMSE) to evaluate the model performance. RRMSE is the root mean squared error normalized by the root mean square value where each residual is scaled against the actual value [26]. In general, RRMSE less than 10% is considered to be excellent model performance. The proposed model has the RRMSE of 2.82% which very low by standard. These performance matrices attest the predictive capability of the model with a high degree of accuracy, that is 99.3%.

Mean Residuals: The difference between the observed value "y" of the attributable indicator and the predicted value "$\hat{y}$" is called the residual ($\hat{e}$).

$$\text{Residual}\left(\hat{e}\right) = \text{Observed value}\left(\text{WCP}\right) - \text{Predicted value}\left(\widehat{\text{WCP}}\right). \qquad (8)$$

Table 2. Model performance.

Performance Metric	Formula	Score
R^2	$1 - \dfrac{\text{SSE}}{\text{SST}}$	99.3%
R^2_{adj}	$1 - \left[\dfrac{\left(1-R^2\right)\left(n-1\right)}{n-k-1}\right]$	99.2%
RMSE	$\sqrt{\dfrac{\sum_i\left(\text{WCP}_i - \widehat{\text{WCP}_i}\right)^2}{N}}$	2.91
RRMSE	$\sqrt{\dfrac{\frac{1}{N}\sum_i\left(\text{WCP}_i - \widehat{\text{WCP}_i}\right)^2}{\sum_i\left(\widehat{\text{WCP}_i}\right)^2}}$	2.82%
MAPE	$\left(\dfrac{1}{N}\sum_i\left|\dfrac{\text{WCP}_i - \widehat{\text{WCP}_i}}{\text{WCP}_i}\right|\right)\times 100$	1.99%

Both the sum and the mean of the residuals should be equal to zero assuming that the regression line is actually the line of "best fit". It is found that the sum and the mean of the residuals are $e^{-17} \approx 0$ and $e^{-19} \approx 0$, respectively. Thus, it satisfies the regression assumption.

Normality of Residuals: One of the key assumptions to verify of the proposed model is the normality of the residuals. The p-value from normality test using Shapiro-Wilk and Anderson-Darling tests are 0.3601 and 0.2421, respectively, which indicate that normality assumptions of the residuals are satisfied at 5% level of significance.

Autocorrelation: Multiple linear regression assumes that each observation in the dataset is independent, that is, no autocorrelation. The degree of correlation of the same indicator between two successive time intervals is defined as the autocorrelation [27]. In financial time series, there is high chance that the next value of the indicator in a series can highly be influenced by its own lagged value. Therefore, it is crucial to test for the autocorrelation of the historical weekly closing prices to identify to what extent the price change is merely a pattern or caused by other factors.

We use the Durbin-Watson statistic to check the autocorrelation in multiple regression for time series data. The Durbin-Watson test statistic ranges from 0 to 4. The test statistic close around 2 means a very low level of autocorrelation where as closer to 0 suggests a stronger positive autocorrelation, and closer to 4 suggests a stronger negative autocorrelation.

The Durbin-Watson test statistic is 1.966 which is close to 2 with p-value = 0.084 > 0.05 for our regression model that estimates WCP^T of the stock, suggests the fact that each observation in the dataset is independent. That is, there is no autocorrelation present in the dataset.

Homoscedasticity: One of the main assumptions for the regression model is the homogeneity of the variance of the residuals.

Figure 5 below, shows that there is no definite pattern of the residuals and they are approximately equally distributed on either sides of reference line. This indicates that the proposed statistical model satisfies the assumption of constant variance of the residuals.

k-Fold Cross Validation: We performed k-fold repeated cross validation on training dataset to evaluate the goodness-of-fit of the proposed model such that the test error gives an idea about the predictive consistency of the analytical model on a new dataset. We used k = 10, that is 10-fold cross-validation, is a re-sampling technique that randomly divides the training data into 10 groups/folds of approximately equal size. The model is fit on 9 (*i.e.* k-1) folds and then the remaining fold is used to compute model performance. This procedure is repeated 10 times; each time, a different fold is treated as the validation set. If the current selected model has a good predicted power, then the mean square error for the training data should be approximately equal to the mean square predictive error of the test data.
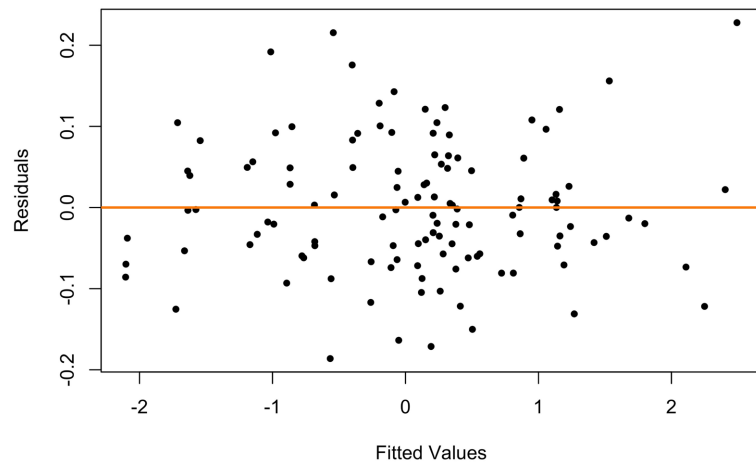
Figure 5. Residuals vs. Fitted values.

In our proposed analytical model, the Mean Squared Error for training data (MSETr) and Mean Squared Predictive Error on the test data (MSPE) are 0.0092 and 0.0088, respectively. The MSETr and MSPE are approximately equal, therefore, we are confident that the proposed analytical model does not suffer from the issues of overfit or underfit, rather it is the best fit highly accurate predictive model.

4. Usefulness of the Proposed Predictive Model

Equation (7) in Subsection 2.4, is the best preferred analytical model to predict the WCP of the MSFT stock of Information Technology Sector Index of S&P 500 and it has its own significant importance in the field of applied finance and statistics. In this section, we illustrate five important usefulness of the proposed model.

1) It identifies the most significant indicators that drive the WCP of the MSFT stock.

In Equation (5), we see that the proposed model identifies all financial and economic indicators that are statistically significantly contribute to the WCP of the stock. We noticed that Beta, P/E Ratio, PEG/Ratio, PSR and GDP are the most statistically contributable indicators to the WCP of the stock where as the FCF/Share, P/B Ratio, Dividend Yield, Interest Rate and ICS have no statistically significant contribution at 5% level of significance.

2) It also identifies the important interactions of the indicators that significantly contribute to the WCP of the stock.

There are 11 statistically significant interactions of the indicators which are identified by the proposed model. We observe that the interaction between Beta and P/B Ratio, P/B Ratio and Dividend Yield, P/B Ratio and P/E Ratio, P/E Ratio and Interest Rate, Dividend Yield and Interest Rate, ICS and PSR, PEG Ratio and Dividend Yield, Beta and PEG Ratio, ICS and GDP, Interest Rate and ICS and, P/B Ratio and Interest Rate statistically significantly contribute to the WCP of the stock.

It is interesting to note that indicators that are not individually statistically significant in the model, their interaction with other indicators seem to be the most significant attributable entities to estimate the WCP of the stock. For instance, P/B Ratio is non-statistically significant indicator in the model. However, its interaction with Beta, P/E Ratio and Dividend Yield are highly significant entities of our proposed model. The interaction of the P/B Ratio with Interest Rate is moderately significant at 5% level of significance.

3) We rank the indicators and their interactions based on the percentage of contribution to the WCP of the stock as shown in Table 3, below.

We see that GDP contributes the highest percentage 14.88% to the WCP of the stock followed by the interaction between Beta and P/B Ratio, 7.53%. Our findings from this study about the GDP, as the most influential indicator to determine WCP of the stock does not support the argument made by Duda [10]. He reported that there is no direct connection between stock market growth and the GDP growth. However, we found that the US GDP is the number one contributing attributable indicators among the above mentioned economic and financial indicators, and it is highly positively correlated with WCP of the stock. It

Table 3. Rank of the most significant indicators and their interactions based on the percentage of contribution to the WCP of the MSFT stock.

Rank	Indicators & Interactions	Contribution (%)
1	GDP	14.88
2	Beta ∩ P/B Ratio	7.53
3	Beta	7.06
4	P/B Ratio ∩ Div Yield	6.60
5	P/B Ratio ∩ P/E Ratio	6.55
6	P/E Ratio ∩ Int R	5.92
7	Div Yield ∩ Int R	5.90
8	P/E Ratio	5.30
9	ICS ∩ PSR	5.25
10	PEG Ratio ∩ Div Yield	5.23
11	PSR	4.62
12	Beta ∩ PEG Ratio	4.61
13	PEG Ratio	4.23
14	ICS ∩ GDP	4.17
15	Int R ∩ ICS	3.24
16	P/B Ratio ∩ Int R	3.22
17	Div Yield	2.29
18	ICS	1.73
19	Int R	1.19
20	P/B Ratio	0.49

makes sense with the fact that growing GDP implies growth in per capita income and thus, investor will have more money to spend/invest on stocks. This enhances economic activities and stimulates stock market as well. Similarly, it is interesting insight to note that the P/B is the least contributing indicator (0.40%) among the list of attributable indicators included in this study. However, its interaction with Dividend Yield, P/E Ratio and Interest Rate in total contributes 16.37% to the WCP of the stock. Therefore, it is very important for the investors, financial analysts and companies to understand the behaviour of each individual indicator to foresee the future performance of the company.

Microsoft is being one of the best companies in terms of dividend policy and it has constantly distributed dividend to its shareholders in its long history. People do invest in stock by evaluating company's dividend policy. The results from our analysis also supports this fact as Dividend Yield, and its interaction with P/B Ratio, Interest Rate and PEG Ratio are found to be statistically significant contributor to estimate the WCP of the MSFT stock.

In our study, we found that PEG Ratio, and its interaction with Beta and Dividend Yield are powerful attributable indicators to estimate the WCP of the stock. Our findings from this study about ICS as a significant attributable indicator to estimate WCP of the stock price also supports the argument made by Lemmon *et al.* (2006) that investors' confidence level exhibits forecasting power for the stock returns [8]. Similarly, our studies show that PSR is one of most contributable individual indicator and its interaction with ICS is also statistically significant entities to predict the WCP of the stock price. This finding about PSR also favors the argument that current saving rates influences future consumption and supports investments [9].

The ranking of indicators included in the model based on the percentage of contribution to the WCP is extremely important in the sense that it serves as a prior knowledge for scientific researchers, business analysts and investors. Having prior knowledge of the strength of attributable indicators and their interactions that significantly contribute to the response can be beneficial in decision making process. It also updates intuitive knowledge in the mind of consumers and service providers so that they can focus more on those influential indicators while making important business decisions.

4) For any given set of indicators, the model predicts the change on WCP of the stock accurately.

In Table 4 below, we can clearly see that the predicted values are very close to the observed values and, thus attest to the accuracy of our proposed model's predictive power of 99.3%. The proposed analytical model can be helpful for the researchers, economists and financial analysts to understand how the WCP of the index varies when any one of the attributable indicators is varied, keeping the other indicators fixed. In other words, understanding the behaviour of the attributable indicators and their interactions help predict the change in the WCP of the stock.

Table 4. The list of observed and predicted values of the AWCP.

Obs.	Observed	Predicted
6	64.00	64.33
12	64.98	64.82
15	66.40	65.01
23	71.21	69.55
24	68.93	70.59
27	73.79	72.68
28	73.04	73.96
39	77.49	78.67
42	84.14	86.47
46	84.26	82.36
49	85.51	84.09
65	93.08	98.86
69	97.70	97.78
74	100.13	101.34
75	100.41	99.81

Obs.	Observed	Predicted
81	108.04	109.38
82	107.58	111.32
103	103.17	100.79
106	108.22	106.80
111	117.05	117.61
112	117.94	111.80
114	120.95	123.42
121	132.45	130.69
123	133.96	133.55
125	138.90	135.95
131	137.39	137.08
133	139.10	136.74
135	139.44	136.51
137	137.73	137.25
146	151.75	155.33

5) Having such an excellent model, the proposed procedure and methodology can be effectively used to develop predictive models for other companies and their business sectors of the S&P 500 to predict their price, thereby facilitating companies, business analysts and investors to make effective financial decisions.

5. Understanding Behavior of Financial & Economic Indicators on the MSFT Stock and the Information Technology Sector Index Price of S&P 500

Research has been conducted to evaluate the effect of financial and economic indicators to the weekly closing price of the Information Technology Sector Index of S&P 500 [11]. Since MSFT is being one of the major players of the index, in this section we compare the effect of significant indicators and their interactions on the WCP through respective predictive models.

Table 5 below, shows the relative importance of the indicators and their interactions based on the percentage contribution to the WCP of the Information Technology Sector Index of S&P 500 [11]. We can see that GDP contributes more than 18% while explaining the variation on the WCP of the index.

Form Section 3.2, Table 3: Rank of the Most Significant Indicators and their Interactions based on the Percentage of Contribution to the AWCP of MSFT stock, we inspect that GDP is the number one most statistically significant individual indicators for explaining the variation on the AWCP of the MSFT stock. It contributes 14.88% to the AWCP of the MSFT stock based on our current proposed model.

Table 5. Rank of the most significant indicators and their Interactions based on the percentage of contribution to the WCP of the information technology sector index.

Rank	Indicators/Interaction	Contributions %
1	GDP	18.60
2	PEG Ratio ∩ Int R	14.23
3	PEG Ratio ∩ GDP	11.65
4	P/B Ratio	7.32
5	PEG Ratio	7.06
6	Beta ∩ P/B Ratio	7.00
7	P/B Ratio ∩ Int R	6.42
8	FCF/Share ∩ GDP	4.29
9	ICS ∩ PSR	3.93
10	Beta ∩ PSR	3.56
11	P/E Ratio	3.40
12	P/E Ratio ∩ Int R	2.90
13	FCF/Share	2.89
14	FCF/Share ∩ DIV Y	2.89
15	PSR	1.71
16	ICS	1.12
17	Div Y	0.54
18	Int R	0.49
19	Beta	0.37

It is important to note that GDP which is a measure of the monetary value of goods and services produced within a country's borders in a given time period, is the most statistically significant attributable entity that positively affects the AWCP of individual stock as well the overall index. Our findings from this study does not support the claim made by Ashutosh Duda [10]. The effect of GDP on the WCP of IT Index is much higher than the MSFT stock. Growing economic activities stimulates overall economic health of the country, thereby increasing a country's GDP. It creates more job opportunities, fosters investment environment and thus, it affects stock market as well.

In this comparative study, we found that the impact of P/B Ratio on the WCP of the MSFT stock and the IT index is different. P/B Ratio has minimum contributing effect on WCP of MSFT stock and ranked on the bottom of Table 3. However, it is ranked number 3 based on percentage contribution to the WCP of IT Index. But, the significant importance of the P/B Ratio on explaining the variation on the WCP is inspected via both predictive models as its interactions with other indicators are statistically significant. The PEG Ratio and its interactions have greater impact on the WCP of the IT Index as compared to PE Ratio. Alternatively, P/E Ratio is found to be more statistically significant indicators while explaining variation on the WCP for MSFT stock. Both ICS and PSR are statistically significant entities of the proposed models and their capabilities for predicting stock/index price can not be undermined. It is interesting to note that Dividend Yield has greater impact on the MSFT stock price as compared WCP of the Information Technology Sector Index of S&P 500.

6. Conclusions

The proposed real data-driven analytical model has its own significance in the field of finance and economics. It is developed using strong theoretical understanding of the statistical concepts and financial domain knowledge. The predictive model satisfies all the statistical assumptions, has been tested and validated and, ensures the predictive accuracy of 99.3%. The highlights of the usefulness of the model presented in the Section 4 are the testimony of the quality of the proposed model and its unique contribution to the field of applied finance and economics.

The paper presents some intriguing findings about the attributable indicators and their interactions that influence the Weekly Closing Price of the stock. For instance, GDP and the Beta are ranked No. 1 and No. 3, respectively, based on the percentage of contribution to the Weekly Closing Price of the index. Other interesting finding of our study is that the interaction effect of the P/B Ratio with Beta, Dividend Yield, P/E Ratio and Interest Rate contributes closely 24% on the stock weekly closing price. However, P/B Ratio itself has minimum impact to the weekly closing price of the stock. Thus, it is important to note that ignoring P/B Ratio can be detrimental while estimating the weekly closing price of the stock. Similarly, P/E Ratio and PEG Ratio are two important financial indicators that

need to be considered. Interest Rate, particularly its interaction with other indicators, ICS and PSR are observed to be powerful economic indicators for stock price prediction. Much research has been conducted in the past to understand the impact of financial and economic indicators on stock returns, however the study of the interactions effect of those indicators on the stock/index price is hardly been explored. In the paper, we present most contributing attributable financial and economic indicators and their interaction effect and discuss the relative importance of those indicators while predicting the Weekly Closing Price of the MSFT stock. The comparative study of indicators between MSFT stock and the Information Technology Sector Index of S&P 500 provides useful insights for investment decisions.

The proposed model is very useful for individual investors and institutions to assess the short and long-term investment strategies. It is also equally important for the company's managers and shareholders to build policies and strategies to keep up the momentum of the stock price by closely monitoring the key indicators and their interaction effect.

The proposed model building procedure and methodology can be effectively used to develop predictive models for individual companies and other business sectors of S&P 500 and we will continue our effort to explore this idea in our future research.

7. Statements and Declarations

To the best of our knowledge, this manuscript represents our original work and meets the ethical standards set by the Committee on Publication Ethics (COPE).

Patent

This research may result in possible patent and is protected by the TTO of the University of South Florida.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Ghysels, E., Santa-Clara, P. and Valkanov, R. (2005) There Is a Risk-Return Trade-Off after All. *Journal of Financial Economics*, **76**, 509-548.
https://doi.org/10.1016/j.jfineco.2004.03.008

[2] Sava, J.A. (2022) Tech GDP as a Percent of Total GDP in the U.S.
https://www.statista.com/statistics/1239480/united-states-leading-states-by-tech-contribution-to-gross-product

[3] Lazy Portfolio ETF (2022, September 30) S&P 500 Sector Returns.
http://www.lazyportfolioetf.com/sp-500-sector-returns

[4] Tang, G.Y. and Shum, W.C. (2003) The Relationships between Unsystematic Risk, Skewness and Stock Returns during up and down Markets. *International Business*

Review, **12**, 523-541. https://doi.org/10.1016/S0969-5931(03)00074-X

[5] Enriquez-Savery, S. (2016) Statistical Analysis of a Risk Factor in Finance and Environmental Models for Belize. USF Tampa Graduate Theses and Dissertations. https://digitalcommons.usf.edu/etd/6231

[6] Lajevardi, S. (2014) A Study on the Effect of P/E and PEG Ratios on Stock Returns: Evidence from Tehran Stock Exchange. *Management Science Letters*, **4**, 1401-1410. https://doi.org/10.5267/j.msl.2014.6.029

[7] Alam, M.D. and Uddin, G. (2009) Relationship between Interest Rate and Stock Price: Empirical Evidence from Developed and Developing Countries. *International Journal of Business and Management*, **4**, 43-51. https://doi.org/10.5539/ijbm.v4n3p43

[8] Lemmon, M. and Portniaguina, E. (2006) Consumer Confidence and Asset Prices: Some Empirical Evidence. *The Review of Financial Studies*, **19**, 1499-1529. https://doi.org/10.1093/rfs/hhj038

[9] Garner, C.A. (2006) Should the Decline in the Personal Saving Rate Be a Cause for Concern? *Economic Review, Federal Reserve Bank of Kansas City*, **91**, 5-28.

[10] Duda, A. (2020) A Study of Interlinkage between Stock Market and GDP Growth. *European Journal of Molecular & Clinical Medicine*, **7**, 3860-3866.

[11] Pokharel, J.K., Tetteh-Bator, E. and Tsokos, C.P. (2022) A Real Data-Driven Analytical Model to Predict Information Technology Sector Index Price of S&P 500. https://arxiv.org/abs/2209.10720

[12] Fiorillo, S. (2021, April 2) Apple and Microsoft: An In-Depth Comparison across 17 Categories. https://seekingalpha.com/article/4417318-apple-microsoft-in-depth-comparison-across-17-categories

[13] Akono, H. (2016) Free Cash Flow and Executive Compensation. *International Journal of Business and Social Science*, **7**, 11-34.

[14] Indrayono, Y. (2019) Predicting Returns with Financial Ratios: Evidence from Indonesian Stock Exchange. *Management Science Letters*, **9**, 1908-1908. https://doi.org/10.5267/j.msl.2019.6.003

[15] Shen, P. (2000) The P/E Ratio and Stock Market Performance. *Economic Review, Federal Reserve Bank of Kansas City*, **85**, 23-36.

[16] Meher, B. K. and Sharma, S. (2015) Is PEG Ratio a Better Tool for Valuing the Companies as Compared to P/E Ratio? A Case Study on Selected Automobile Companies. *International Journal of Banking, Risk and Insurance*, **3**, 48-52. https://doi.org/10.21863/ijbri/2015.3.2.012

[17] Chen, J. (2022, September 22) Federal Funds Rate: What It Is, How It's Determined, and Why It's Important. https://www.investopedia.com/terms/f/federalfundsrate.asp

[18] University of Michigan (2022, June 22) University of Michigan: Consumer Sentiment [UMCSENT]. FRED, Federal Reserve Bank of St. Louis, St. Louis. https://fred.stlouisfed.org/series/UMCSENT

[19] U.S. Bureau of Economic Analysis (2022, June 20) Personal Saving Rate [PSAVERT]. FRED, Federal Reserve Bank of St. Louis, St. Louis. https://fred.stlouisfed.org/series/PSAVERT

[20] Fernando, J. (2022, September 29) Gross Domestic Product (GDP). https://www.investopedia.com/terms/g/gdp.asp

[21] Chou, Y.M., Polansky, A.M. and Mason, R.L. (1998) Transforming Non-Normal

Data to Normality in Statistical Process Control. *Journal of Quality Technology*, **30**, 133-141. https://doi.org/10.1080/00224065.1998.11979832

[22] Tibshirani, R. (1996) Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, **58**, 267-288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x

[23] Moorthi, A. (2020, November 26) How Lasso Regression Works in Machine Learning. Dataaspirant https://dataaspirant.com/lasso-regression

[24] Oyeyemi, G.M., Ogunjobi, E.O. and Folorunsho, A.I. (2015) On Performance of Shrinkage Methods—A Monte Carlo Study. *International Journal of Statistics and Applications*, **5**, 72-76.

[25] Guyon, I. (2008) Practical Feature Selection: From Correlation to Causality. In: *Mining Massive Data Sets for Security, Advances in Data Mining, Search, Social Networks and Text Mining, and Their Applications to Security*, IOS Press, Dordrecht, 27-43.

[26] Padhma, M. (2022, September 22) End-to-End Introduction to Evaluating Regression Models. https://www.analyticsvidhya.com/blog/2021/10/evaluation-metric-for-regression-models

[27] CFI Team (2022, May 6) Autocorrelation. https://corporatefinanceinstitute.com/resources/knowledge/other/autocorrelation