Trading Volume as a Predictor of Market Movement

1 author:

Edson Kambeu
BA ISAGO University
6 PUBLICATIONS   **5** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Project   Financial Economics View project

Finance & Banking Studies

Trading volume as a predictor of market movement: An application of Logistic regression in the R environment

Edson Kambeu

Corresponding Author: Department of Business Management , BAISAGO University, Francistown, Botswana. +267 241 8780

ORCID ID: https://orcid.org/0000-0001-9166-255X

Abstract

 A Logistic regression model become a popular model because of its ability to predict, classify and draw relationships between a dichotomous independent variable and dependent variables. On the other hand, the R programming language has become a popular language for building and implementing predictive analytics models. In this paper, we apply a logistic regression model in the R environment in order to examine whether daily trading volume at the Botswana Stock Exchange influence daily stock market movement. Specifically, we use a logistic regression model to find the relationship between daily stock movement and the trading volumes experienced in the recent five previous trading days. Our results show that only the trading volume for the third previous day influence current stock market index movement. Overall, trading volumes of the past five days were found not have an impact on today's stock market movement. The results can be used as a basis for building a predictive model that utilizes trading as a predictor of stock market movement.

Keywords: Stock exchange, Market movement, Logistic regression, R programming

JEL classification: G17; G21

Introduction

For years, traders and financial analysts have been preoccupied with predicting stock market behavior. Predicting stock market movement has been one of the problems of interest in the field of finance (Pagolu et al, 2016). It is essential to financial analysts and traders because it allows them to formulate better trading strategies. Moreover, it allows traders to hedge against market risks and profit from possible arbitrage opportunities (Alrasheedi, 2012; Song,2016). However, predicting stock market behavior is not an easy task. This is because of the dynamic and complex nature of the stock market, which makes making reliable predictions challenging (Kim and Sayama, 2017; Huang et al, 2005). The challenge in financial literature has been identifying the most effective predictors of stock market movement.

Most traditional studies on stock market prediction use macroeconomic fundamental factors such as Gross Domestic Product and CPI to predict market direction (Wang, 2014; Rajput and Bobde,2016). Trading volume has also emerged as one of the useful stock market variables that can be used to predict stock market movement. This is because trading volume is a quantification of supply and demand, which in turn determines the dynamics of equilibrium prices (Remorov, 2014). In other words, trading volume tends to influence the dynamics of stock price movements. Being a determinant of stock price movement, past trading volume can indeed potentially be used to predict future stock price movements (Habib, 2011).

According to the Efficient market Hypothesis, stock prices change as a result of information flow. To be an effective predictor of stock price movement, trading volume should therefore contain useful information that can influence trading decisions. In other words, trading volume should be a proxy for information flow which eventually influence stock returns. Brooks (1998) argued that trading volume contain useful information that can be used to model financial variables of interest such as volatility and returns. According to Hussain et al (2014) trading volume acts as a signal to unformed investors about what is going to happen to a stock. This means that trading volume can indeed convey useful information to forecast stock price movements.

There are several studies that support the viewpoint that past trading volume can be a predictor of future stock prices. For example, a study by Ciner (2003) concluded that past trading volume can be used to predict returns. Another example is a study by Choi et al (2013) which found trading volume to have a strong predictive power for the financial markets of Japan and Korea. Also, a study by Tehranchian et al (2014) found a positive and bi-directional relationship between stock returns and volume in Tehran stock market. Thus, there has been in interest in building models that utilize trading volume as a predictor of stock market returns.

Predicting stock market behavior has always relied on traditional statistical models. However, in recent times there has however been a diversion from the use of standard statistical techniques to machine learning techniques in predicting stock market prices and movements. Machine learning refers to computer algorithms that learn from past data. The rise of machine learning techniques in solving predictive problems in finance has primarily been due to the increased computational power and availability of big financial data.

There are several examples of researchers who have utilized machine learning techniques to predict stock market movements. Qiu and Song (2016) built a model that predicted the direction of the next day's price of the Japanese stock market index by using an optimized artificial neural network (ANN) model. Grigoryan (2015) also used applied Artificial Neural Network on the TAL1T stock of Nasdaq OMX Baltic stock exchange to predict stock prices. Khaidem et al (2016) used a machine learning algorithm called random forest to predict the direction of future prices. Dutta et al (2012) predicted stock performance in the Indian stock market using logistic regression, a supervised machine learning algorithm.

We are also interested in applying machine learning techniques to predict stock market movement at the Botswana Stock Exchange using trading volume. However, building that predictive model requires one to firstly find out if trading volume can efficiently predict stock market movement. In this study, we use a logistic regression model to investigate the predictability of stock market movement using trading volume at the Botswana Stock Exchange. Logistic regression is a popular statistical learning model, chosen because of its ability to draw a relationship between a dichotomous dependent random variable (in this case stock market movement) and its independent variables. To enable reproducibility of the study, we apply the logistic regression model using R programming language. R is a popular statistical programming language widely used by statisticians to analyze big data and implement predictive analytics. In the next sections, we discuss the formulation of the logistic regression model and the implementation of the model in R to find out if trading volume can be used as a predictor of stock market movements.

Model-Logistic Regression Model

The stock market index can move up or down on a particular day. At time t , stock index movement takes the value 1 if there an upward movement in the index or 0 if there is a downward movement in the stock index.

$$y_t = \begin{cases} 1 \ upward \ movement \\ 0 \ \ down \ movement \end{cases}$$

We assume that that today's stock index movement can be predicted by trading volumes for the past five days. The predictors can be represented as a linear sum as follows:

$$z = \beta_0 + \beta_1 X_{t-1} + \beta_2 X_{t-2} + \beta_3 X_{t-3} + \beta_4 X_{t-4} + \beta_5 X_{t-5}$$

The variables $X_{t-1}, X_{t-2}, X_{t-3}, X_{t-4}$ and X_{t-5} represent trading volume for past 1 day, 2 days, 3 days, 4 days and 5 days from time t. The probability of an upward movement at time t in the stock index given the predictors z can be written as a conditional probability as follows:

$$P(Y_t) = P(y_t = 1/z)$$

Applying the logistic model function, the probability of an upward movement at time t is then written as follows.

$$P(Y_t) = \frac{1}{1+e^{-z}}$$

We express the probability of an upward movement as an odds ratio. This is shown mathematically as follows:

$$\frac{P(Y)}{1 - P(Y)} = \frac{\frac{1}{1 + e^{-z}}}{1 - \frac{1}{1 + e^{-z}}}$$

$$= e^z$$

The odds ratio equation is then log-transformed as follows.

$$\log\left[\frac{P(Y_t)}{1 - P(Y_t)}\right] = \log(e^z)$$

Letting the odds ratio be equal to Y_t :

$$Y_t = \frac{P(Y_t)}{1 - P(Y_t)}$$

The log-transformed equation will simplify to:

$$logit \ Y_t = z$$

Substituting z for the sum of linear predictors, the logistic regression model will be specified as follows:

$$logit\ \boldsymbol{Y_t} = \beta_0 + \beta_1 X_{t-1} + \beta_2 X_{t-2} + \beta_3 X_{t-3} + \beta_4 X_{t-4} + \beta_5 X_{t-5}$$

Y_t is the log odds of an upward stock price movement given the predictors $X_{t-1}, X_{t-2}, X_{t-3}, X_{t-4}$ and X_{t-5} . The above logistic regression model is used to find out if trading volume can be a predictor of stock price movement.

Data analysis and results

The daily data for the Botswana Domestic Company Index from 2015 to 2017 was collected and analysed using the R programming language. To make the data analysis reproducible, the entire steps used to analyze the data and apply the logistic are discussed in this section.

- **Importing data into R**

We began the analysis by importing the data into R. The read_excel() function provided by the readxl package was used import the excel data into R.

```
> library(readxl)
> bsedata <- read_excel("bsedata.xlsx")
```

- **Exploring the structure of the data frame**

We explored the data frame by firstly viewing the first six observations of the data frame using the head () function.

```
> head(bsedata)
# A tibble: 6 x 3
  year_month  volume      dci
       <chr>   <chr>    <chr>
1   Jan_2014  151853 9064.31
2   Jan_2014   97069 9068.48
3   Jan_2014  247955 9078.46
4   Jan_2014 5313437 9084.03
```

The bse data has three variables y_month, volume and dci. The variable year_month represents the month and year. The volume variable represents the daily trading volume and dci represents the domestic companies index. The function str () was called to view the structure of the data frame.

```
> str(bsedata)
Classes 'tbl_df', 'tbl' and 'data.frame':    742 obs. of  3 variables
:
 $ year_month: chr  "Jan_2014" "Jan_2014" "Jan_2014" "Jan_2014" ...
 $ volume    : chr  151853 97069 247955 5313437 164632 ...
 $ dci       : chr  9064 9068 9078 9084 9093 ...
```

The variables dci and volume are being shown as character variables instead of numeric variables. And as such we converted the character variables into numeric variables using the as.numeric() function.

```
>bsedata$volume <- as.numeric(bsedata$volume)
> bsedata$dci <- as.numeric(bsedata$dci)
```

- **Identifying missing values**

A logistic regression model should not be applied when there are missing values in the data. We therefore explored the data to find out if there are any missing values using the summary() function.

```
> summary(bsedata)
  year_month          volume              dci
 Length:742       Min.   :        0   Min.   : 8835
 Class :character  1st Qu.:   193422   1st Qu.: 9441
 Mode  :character  Median :   698802   Median : 9711
                   Mean   :  2819482   Mean   : 9866
                   3rd Qu.:  2785713   3rd Qu.:10367
                   Max.   :159425347   Max.   :11097
```

The summary results show that there were no missing values in any of the three variables of the data frame.

- **Identifying outliers**

A valid logistic regression model should not have outliers. The next step was therefore to identify if there were any outliers in the data. We used a box plot to visualize the outliers in data. The observations beyond the boxplot whiskers are regarded as outliers. The following codes were used to plot the boxplots for the two variables.

```
> boxplot(bsedata$dci, xlab ="Domestic Company Index(DCI)", main = "Boxplo
t:DCI")
> boxplot(bsedata$volume, xlab = "Trading Volume", main = "Boxplot:Trading
volume")
```

The resultant box plots are shown below:

Boxplot:DCI

Domestic Company Index(DCI)

Boxplot:Trading volume

Trading Volume
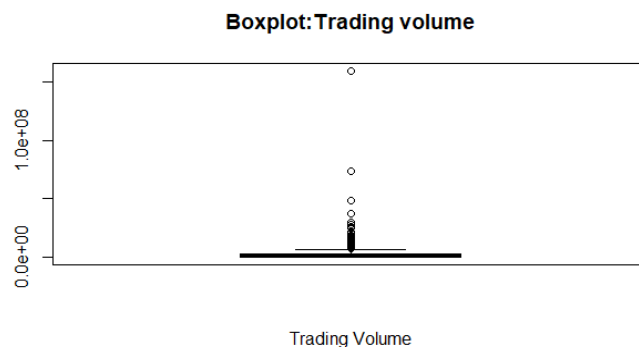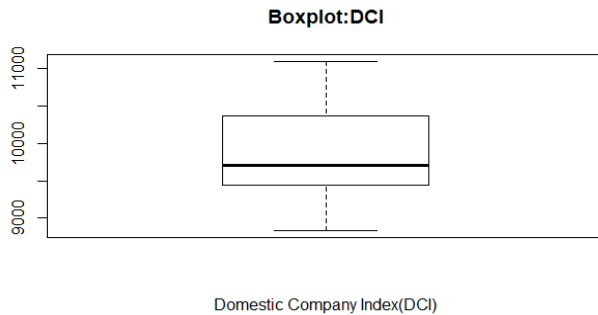
The box plots showed that there were no outliers in the dci variable, but there several outliers for the volume variable. We then decide to identify the actual values in the volume variable which were regarded as outliers. We therefore computed the box plot stats and printed the actual values which were beyond the whiskers.

```
> index_outliers<- which(bsedata$volume %in% boxplot(bsedata$volume)$out)
> length(index_outliers)
[1] 80
```

The series of codes above showed that the variable volume had 80 outlier observations. The next step was to identify actual outlier values to find out if the outliers were errors. Because they were many outlier values, we decided to have a glimpse of the first 6 outlier values using the following code.

```
> y <-boxplot(bsedata$volume)$out
> head(y)
[1] 13587494 10090423  8956583 73559302  7744877 21307758
```

The values did not seem to be errors. However, there was need to check the original data to be certain. We therefore identified the actual positions of the first six outlier values by calling the head function on the index of outliers. This was achieved using the following code.

Page62

```
head(index_outliers)
[1] 21 23 48 50 59 76
```

The code revealed that the first 6 outlier values were observation 21, 23, 48, 50 and 76 in the data. Using the positions of the observations, we rechecked the original data source to find out if they were errors in data entry. We found out that the values were not a result of errors and as such we decided not to delete them from the data. Rather, we decided to later check their influence on the logistic regression model.

- **To create a dichotomous/binary variable**

The response variable for our logistic model is the stock market index movement. The index movement for a particular day is said to be "up " if the index for the current day is higher than for the previous day. Conversely, it is said to be "down" if the index for the current day is higher than the previous day index. We used a for loop that loops over the column dci to create the column "movement".

```
> for(i in 1: length(bsedata$dci)){
+   if(bsedata$dci[i+1] > bsedata$dci[i]){
+     bsedata$movement[i+1] <- "up"
+   } else{
+     bsedata$movement[i+1] <- "down"
+   }
+ }
Error in if (bsedata$dci[i + 1] > bsedata$dci[i]) { :
  missing value where TRUE/FALSE needed
> head(bsedata$movement)
[1] NA   "up" "up" "up" "up" "up"
```

- **Creating the response variable for the logistic regression model**

To be able to apply the logistic regression model using our data, we were supposed to convert the response variable "movement" to a numeric variable in which "1" represents upward movement and 0 represents downward movement. We also used a for loop to create a new column named "movement1" which assigns 1 to a variable if today's index is greater than previous day's index and "0" otherwise. The following code was used:

```
> bsedata$movement1 <- NA
> for(i in 1: length(bsedata$dci)){
+   if(bsedata$dci[i+1] > bsedata$dci[i]){
+     bsedata$movement1[i+1] <- 1
+   } else{
+     bsedata$movement1[i+1] <- 0
+   }
+ }
Error in if (bsedata$dci[i + 1] > bsedata$dci[i]) { :
  missing value where TRUE/FALSE needed
> head(bsedata$movement1)
[1] NA  1  1  1  1  1
```

- **Creating predictor variables of the logistic model**

The predictors for the logistic model are the trading volumes of the last five previous trading days. As such we created new columns of the lagged variable "volume". The mutate function from the dplyr package was used to create the columns. The new data frame is renamed bsedata2. The following code was used to create the columns.

```
> bsedata2 <- bsedata %>% mutate(lagvol_1 = lag(volume),
+                  lagvol_2 = lag(volume,2),
+                  lagvol_3 = lag(volume,3),
+                  lagvol_4 = lag(volume, 4),
+                  lagvol_5 = lag(volume, 5))
> head(bsedata2)
# A tibble: 6 x 10
  year_month  volume     dci movement movement1 lagvol_1 lagvol_2 lagvol_3
       <chr>   <dbl>   <dbl>    <chr>     <dbl>    <dbl>    <dbl>    <dbl>
1   Jan_2014  151853 9064.31    <NA>        NA       NA       NA       NA
2   Jan_2014   97069 9068.48      up         1   151853       NA       NA
3   Jan_2014  247955 9078.46      up         1    97069   151853       NA
4   Jan_2014 5313437 9084.03      up         1   247955    97069   151853
5   Jan_2014  164632 9092.85      up         1  5313437   247955    97069
6   Jan_2014  178996 9104.54      up         1   164632  5313437   247955
# ... with 2 more variables: lagvol_4 <dbl>, lagvol_5 <dbl>
```

In the code above, lagvol_1, lagevol_2, lagvol_3, lag_vol4 and lagvol_5 refers to the trading volumes of the previous 1,2,3,4 and 5 days respectively from today.

- **Checking correlation among the predictors**

Before applying a logistic model on data, it is also important to check if there is multicollinearity among variables. In other words, it is important to check if there are predictor variables that are highly correlated. The select function from the dplyr package was used to select the predictor variables. The base R function cor() was used to compute the correlation matrix.

```
bsedata3 <- select(bsedata2, lagvol_1:lagvol_5)
> cor(bsedata3, use = "pairwise")
          lagvol_1  lagvol_2  lagvol_3  lagvol_4  lagvol_5
lagvol_1  1.00000   0.00500  -0.00076  -0.00412  -0.03717
lagvol_2  0.00500   1.00000   0.00504  -0.00070  -0.00414
lagvol_3 -0.00076   0.00504   1.00000   0.00385  -0.00043
lagvol_4 -0.00412  -0.00071   0.00385   1.00000   0.00428
lagvol_5 -0.03716  -0.00414  -0.00043   0.00427   1.00000
```

The above correlation matrix show that all the predictor variables had very low correlations with each other.

- **The complete Logistic regression model**

With the response binary variable having been created and the predictor variables determined, the following logistic model named "my_logit" was specified. The following code was used to specify the logistic regression model in R.

```
> my_logit <- glm(movement1 ~ lagvol_1 + lagvol_2 + lagvol_3 + lagvol_4 +
lagvol_5, data = bsedata2, family = binomial)
```

In the specification above, "movement1" is the responsible variable in which they are only two binary responses and the lagvol_1, lagvol_2, logvol_3, lagvol_4 and lagvol_5 are the predictor variables representing trading volumes of the last five previous days from today. To obtain the results of the estimated logistic model, the summary function was called.

```
> summary(my_logit)

Call:
glm(formula = movement1 ~ lagvol_1 + lagvol_2 + lagvol_3 + lagvol_4 +
    lagvol_5, family = binomial, data = bsedata2)

Deviance Residuals:
   Min      1Q   Median      3Q     Max
-1.604  -1.224   1.003   1.127   1.312

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  8.836e-02  1.014e-01   0.871   0.3837
lagvol_1    -1.545e-08  1.219e-08  -1.267   0.2051
lagvol_2     5.387e-09  1.009e-08   0.534   0.5934
lagvol_3     2.665e-08  1.559e-08   1.709   0.0875 .
lagvol_4     1.088e-08  1.147e-08   0.949   0.3426
lagvol_5    -2.005e-09  9.528e-09  -0.210   0.8333
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The results showed that lag_vol1, lag_vol2, lagvol_4 and logvo_5 were insignificant predictors of movement1. Only lag_vol3 was a significant predictor of movement1 ate 10% level of significance. This implies that only the trading volume for the past three days from today, influences today's stock market movement. In other words, today's trading volume can be used to predict stock market movement for the third day from today.

We further analysed the overall influence of trading volumes of the past five days on today's stock price movement using the Wald test. The wald.test from the aod package was used to perform the test. The following code was used to perform the Wald Test.

```
> library(aod)
> wald.test(b = coef(my_logit), Sigma = vcov(my_logit), Terms = 2:6)

Wald test:
Chi-squared test:
X2 = 5.7, df = 5, P(> X2) = 0.33
```

The Wald Test revealed that overall trading volume for the past five days from today do not influence the today's market movement.

- **Diagnostic Test: Influence test**

Earlier, we had observed the existence of outliers in our data, specifically for the volume variable. Such outlier observations, if removed can cause can have a significant impact on the coefficient estimates of the logistic regression. The Cook's distance is used to identify the influential observations. The observations with the largest Cook's distance are the most influential observations which can possible significantly change the coefficient estimates of a logistic regression. The influence test is performed using the influencePlot function provided in the car package. The following code was used.

```
> library(car)
> influencePlot(my_logit, col = "red", id.n =3)

      StudRes         Hat      CookD
54  -1.690860 0.13877947 0.07587630
478 -1.639637 0.04179867 0.01989413
609  1.035714 0.50890054 0.13856988
612 -1.470383 0.55161264 0.36626464
```

The resultant influence plot is shown in the figure below.
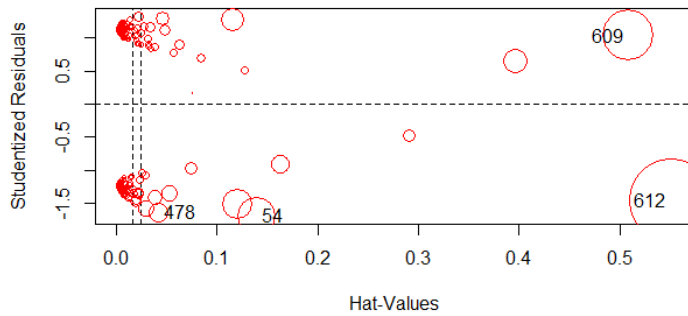
The table and the plot above show that the three observations with the largest influence are observations 54, 609 and 612. We then re-specified another logistic regression model without the most three influential observations and renamed it my_logit3. The coefficients of my_logit are then compared with the coefficients of the re-specified my_logit3. The following series of codes are used to compare the two models.

```
> my_logit3 <- glm(movement1 ~ lagvol_1 + lagvol_2 + lagvol_3 + lagvol_4 +
lagvol_5, data = bsedata2,subset = c(-54, -609,-612), family = binomial)
> compareCoefs(my_logit, my_logit3, pvals = TRUE)
Calls:
1: glm(formula = movement1 ~ lagvol_1 + lagvol_2 + lagvol_3 + lagvol_4 +
   lagvol_5, family = binomial, data = bsedata2)
2: glm(formula = movement1 ~ lagvol_1 + lagvol_2 + lagvol_3 + lagvol_4 +
   lagvol_5, family = binomial, data = bsedata2, subset = c(-54, -609, -612
))

               Model 1    Model 2
(Intercept)    0.0884     0.0592
SE             0.1014     0.1083
Pr(>|z|)        0.384      0.585

lagvol_1     -1.54e-08  -1.48e-08
SE            1.22e-08   1.21e-08
Pr(>|z|)        0.205      0.219

lagvol_2      5.39e-09  -4.05e-09
SE            1.01e-08   1.42e-08
Pr(>|z|)        0.593      0.776

lagvol_3      2.66e-08   2.58e-08
SE            1.56e-08   1.55e-08
Pr(>|z|)        0.087      0.097

lagvol_4      1.09e-08   2.03e-08
SE            1.15e-08   1.54e-08
```

The results of the comparison show that there are no significant differences between coefficients of my_logit and my_logit3. The lagvol_3 also remains significant in both models. This means some of the most influential observations (which are basically the outliers) had no impact on our findings revealed by the my_logit 3 model.

Conclusion

In this paper, we used the logistic regression model to find out if trading volume can be a predictor of stock market movement. Our results revealed only the trading volume for the third most recent day is a significant predictor of today's stock market movement. This implies that trading volume for the third day contains some useful information about the possible future stock market movement. As such today's trading volume can be used to predict the future stock market movement of the third day from today. Overall, our results revealed that trading volume for the past five days do not influence today's stock market movement. This implies that only trading volume of a specific third day from today and not the overall trading volumes of the most recent five days influence today's stock market movement. This study intended to find out if trading volume can be used as a predictor of future stock market movements. The results of this study can be used as a basis for building predictive models that utilizes trading volume as a predictor of stock market movement. Therefore, future studies should investigate focus on formulating powerful predictive models that utilize trading volume as one of the predictor variables.

References

Alrasheedi, M (2012). Predicting Up/Down Direction using Linear Discriminant Analysis and Logit Model: The Case of SABIC Price Index. *Research Journal of Business Management*, 6 (4), 121-133. DOI: 10.3923/rjbm.2012.121.133.

Brooks, C. (1998). Predicting stock index volatility: Can market volume help? *Journal of Forecasting*, 17(1), 59-80. https://doi.org/10.1002/(SICI)1099-131X(199801)17:1<59::AID-FOR676>3.0.CO;2-H

Choi, K.H; Kang, S.H & Yoon, S.M (2013). Relationship between Stock Returns and Trading Volume: Domestic and Cross-Country. Evidence in Asian Stock Markets. *Proceedings of the 2013 International Conference on Economics and Business Administration,* 33-39

Ciner, C (2003). Dynamic Linkages Between Trading Volume and Price Movements: Evidence for Small Firm. *The Journal of Entrepreneurial Finance Stocks*, 8(1), 87-102. https://digitalcommons.pepperdine.edu/jef/vol8/iss1/6

Dutta, A; Bandopadhyay, G & Sengupta, S (2012). Prediction of Stock Performance in the Indian Stock Market Using Logistic Regression. *International Journal of Business and Information*, 7(1), 105-136.

Grigoryan, H (2015). Stock Market Prediction using Artificial Neural Networks. Case Study of TAL1T, Nasdaq OMX Baltic Stock. *Database Systems Journal*, 6(2), 14-23.

Habib, N. M (2011). Trade Volume and Returns in Emerging Stock Markets. An Empirical Study: *The Egyptian Market International Journal of Humanities and Social Science*, 1(19).

Huang, W; Nakamori, Y & Wang, S (2005). Forecasting stock market movement direction with support vector machine. Computers & Operations Research, 32, 2513-2522. http://dx.doi.org/10.1016/j.cor.2004.03.016

Hussain, S; Jamil, H; Javed, M & Ahmed, W (2014). Analysis of Relationship between Stock Return, Trade Volume and Volatility: Evidences from the Banking Sector of Pakistani Market. *European Journal of Business and Management*, 6(20),57-61.

Khaidem L., Saha S. & Dey S.R (2016). Predicting the direction of stock market prices using random forest. *arXiv preprint arXiv:1605.00003*

Kim, M & Sayama, H (2017). Predicting stock market movements using network science: An information theoretic approach. *Applied Network Science* , 2-35. https://doi.org/10.1007/s41109-017-0055-y.

Pagolu, V. S ; Challa, K.N.R & Panda, G(2016).Sentiment Analysis of Twitter Data for Predicting Stock Market Movements. *Proceedings of the International Conference on Signal Processing, Communication,Power and Embedded System, 28 October 2016*

Qiu, M & Song Y (2016). Predicting the Direction of Stock Market Index Movement Using an Optimized Artificial Neural Network Model*. PLoS ONE*, 11(5): e0155133, 1-11. https://doi.org/10.1371/journal. pone.01551

Rajput, V and Bobde, S (2016). Stock market forecasting techniques: Literature survey. *International Journal of Computer Science and Mobile Computing*, 5(6), 500-506.

Remorov, R (2014). Stock Price and Trading Volume during Market Crashes. *International Journal of Marketing Studies*, 6(1), 21-30. http://dx.doi.org/10.5539/ijms.v6n1p21

Tehranchian, A.M; Behravesh, M & Hadinia, S (2014). On the Relationship between Stock Returns and Trading Volume: A Case Study. *European Online Journal of Natural and Social Sciences*, 3(3),425-431.

Wang, Y (2014). Stock price direction prediction by directly using prices data: an empirical study on the KOSPI and HIS. *International Journal of Business Intelligence and Data Mining*, 9(2), 145-160. .https://doi.org/10.1504/IJBIDM.2014.065091