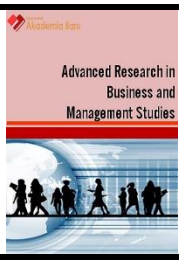
Penerbit Akademia Baru

Journal of Advanced Research in Business and Management Studies

Journal homepage: www.akademiabaru.com/arbms.html
ISSN: 2462-1935



Classify Stock Market Movement Based on Technical Analysis Indicators Using Logistic Regression

Open Access

Bahtiar Jamili Zaini[1,*], Rosnalini Mansor[1,♣], Norhayati Yusof[1,♥], Beh Hui Sang[1]

[1] School of Quantitative Sciences, Universiti Utara Malaysia, 06010 Sintok, Kedah, Malaysia

ARTICLE INFO

Article history:
Received 15 January 2019
Received in revised 26 February 2019
Accepted 13 March 2019
Available online 19 March 2019

Keywords:
Stock market, technical analysis, logistic regression, classification

ABSTRACT

Stock market has been the centre of attraction for investors for a long period of time. The investor's goal is to buy the stock, hold it for a period, and then, sell the stock for more investor paid for it. Many people invest to create wealth and to gain a rich reward. By investing in the stock market, it will improve the returns equity. In this study, the main focus is to predict the future stock price movement for one company listed in Bursa Malaysia. This study used eight months daily basis of historical data to model the relationship using logistic regression. By using logistic regression, stock market movement able to predict the stock price movement, either an increasing trend or unchanged or decreasing movement. Seven technical indicators were used as predictor variables in model formulation, which were Moving Average, Exponential Moving Average, Relative Strength Index, Moving Average Convergence Divergence, Rate of Change, Stochastic Oscillator, and Volume Trading. The results shown that the percentage of correctly classified stock market movement is 86% using in-sample validation data and 71.43% in out-of-sample data. At the end of the model logistic regression formulation, four significant technical indicators to predict the price movement of stock market movement were identified.

Copyright © 2019 PENERBIT AKADEMIA BARU - All rights reserved

1. Introduction

Stock market has been a center of attraction for investors over a long period of time because the stocks generate higher investment returns as compared to other type of investment. Investors always want to know the expectation of investment return before they start to invest. The investor's goal is to buy the stock, hold it for a certain period, and then sell the stock higher than initial price. Most of the investors usually hold the stocks for a long period, however to be successful, they have to buy and sell the stock at the right time for the right price.

* *Corresponding author.*
E-mail address: bahtiar@uum.edu.my (Bahtiar Jamili Zaini)
♣ *Corresponding author.*
E-mail address: rosnalini@uum.edu.my (Rosnalini Mansor)
♥ *Corresponding author.*
E-mail address: norhayati@uum.edu.my (Norhayati Yusof)

The buying and selling of stock price are influenced by investor's sentiment, interest rate or government action. These factors are uneasy to quantify. Therefore, the stock price movements are very hard to predict. Additionally, there are several methods of assessing the stock price. For instance, the methods are neural network [1-3], data mining [4-6], time series forecasting [7], logistic regression [8], genetic algorithm [9-10] and fuzzy logic [11]. Based on Fayek *et. al.* [10], technical indicators are used to generate trading signals of buying or selling. Some of the indicators also can be used to predict stock movements, such as Relative Strength Index (RSI), Moving Average Convergence-Divergence (MACD), Stochastic Oscillators (SO), Rate of Change (ROC), Moving Average (MA), Exponential Moving Average (EMA), closing price and volume trading (VT). By using these technical indicators with the assessing stock price methods, the investor's profit can be improved.

Therefore, this study aims to identify the significant technical indicators to predict stock market movement by using logistic regression. Then, the performance of the logistic model measured using classification rate value. The rest of the paper is organized as follows. Section 2 describe the research methodology. The research analysis and results are summarized and discussed in Section 3. In Section 4, conclusions and future research suggestion of this study are presented.

2. Methodology

2.1 Data and Variables

The data are from Bursa Malaysia which previously known as Kuala Lumpur Stock Exchange (KLSE). There are more than eight hundred companies listed under Bursa Malaysia. This paper only focuses on one company. The data consists of daily data which partition into two parts where data from 4 January 2016 until 30 August 2016 were used to formulate prediction model, and 1 September 2016 until 9 September 2016 to validate the prediction model.

In this study, the stock movement is dependent variable with two categories which defined as follows,

$$Y = \begin{cases} 1, & \text{current closing price is higher than or equal to previous price} \\ 0, & \text{otherwise} \end{cases}$$

where the category one represents the stock movement is going up or remain unchanged based on current closing price. The other category means that the movement is going down.

A total of seven technical indicators are used as the predictor variables. The indicators are MACD, RSI, SO, MA, EMA, ROC and VT. The details about these variables are explained in the following section.

2.1.1 Moving average convergence- divergence

Moving Average Convergence Divergence (MACD) is considered as a leading indicator, but with a bit of lag. The MACD suggest a buy signal when the MACD line crosses above the signal line or the MACD line crosses above the zero line. However, a sell signal is generated when the MACD line crosses below the signal line or crosses down the zero line.

2.1.2 Relatives strength index

Relative Strength Index (RSI) is a leading indicator that measures the speed and the price movement. RSI oscillates between zero and 100. RSI is a kind of momentum oscillators that is used

to calculate the market recent gains against its recent losses and translates that information into a number between 0 and 100 [12]. The value of RSI is considered overbought when above 70 and oversold when below 30.

2.1.3 Stochastic oscillator

The Stochastic Oscillators (SO) is a momentum indicator in the stock market that refers to the current price relative to its price range over a set number of periods [13]. It follows the speed or the momentum of price. Because the SO is range bound, is also useful for identifying overbought and oversold levels. The value of SO is considered overbought when above 80 and oversold when below 20.

2.1.4 Moving average

Moving Average (MA) refers to the simple average price of a stock price over a defined number of time periods. This function returns the value of moving average of a field over a given period of time. The short-term MA average ranges are between 5 to 25 days. MA used to identify the trend direction and to determine support and resistance levels.

2.1.5 Exponential moving average

Exponential Moving Average (EMA) is to reduce the lag by applying more weight to recent prices [9]. When calculating the EMA, the weight or multiplier is not consistent but place more importance on recent price depends on the number of periods in the moving average.

2.1.6 Rate of change (ROC)

Following Fayek *et. al.* [10], ROC as an indicator to measure the percentage increase or decrease in stock price over a given period of time. ROC give a signal trend is upward, if ROC is greater than zero. Otherwise, ROC gives a downtrend signal if the ROC is less than zero.

2.1.7 Volume trading (VT)

VT is an indicator that demonstrates the amount of financial instrument traded over a certain time period. This indicator is not based on stock price. High volume points to a high interest in an instrument at its current price and vice versa. If the stock price is shows increasing but the volume is going down it means that there fewer investor buying at a higher price. Therefore, it can give a signal that potential change in direction of the stock price.

2.2 Logistic Regression Model

Logistic regression is commonly method used to predict dependent variable as either success or failure based on predictor variables. The advantage of this method is it does not need assumption of distribution for the error term. The logistic model or logit model for this study is written as follows,

$$logit\ (p) = ln\left(\frac{p}{1-p}\right) = b_0 + b_1\ MACD + b_2\ RSI + b_3\ SO + b_4\ MA + b_5\ EMA + b_6\ ROC + b_7\ VT$$

where *p* is the probability of stock movement is going up or remain unchanged, and $\frac{p}{1-p}$ is the odds of stock movement is going up or remain unchanged. The odds also can be written as follows,

$$odds = \frac{p}{1-p} = e^{b_0} \times e^{b_1 MACD} \times e^{b_2 RSI} \times e^{b_3 SO} \times e^{b_4 MA} \times e^{b_5 EMA} \times e^{b_6 ROC} \times e^{b_7 VT}$$

where b_i is regression coefficient estimated by using maximum likelihood estimation, and e^{b_i} is the odds ratio (OR) for the predictor variable *i*. OR value gives the relative amount of the odds will increase or decrease when the predictor variable increased by one unit.

The analysis started with logit model formulation by using the first part of data partitioned. The significant predictors are determined by Wald statistic. The statistic is written as follows,

$$W_i = \left(\frac{b_i}{se_{b_i}}\right)^2 \quad i = 1,2,\ldots,7$$

where b_i is the regression coefficient for predictor *i*, and se_{b_i} is the standard error of regression coefficient for predictor *i*. This statistic distributed as Chi-square distribution.

Then, the model is reformulated based on the important predictors only. Subsequently, the new model is applied to the second part of data partitioned for predicting the stock movement category. The performance of the model is measured by the percentage of correctly classified the category. This value known as classification rate.

3. Analysis and Results

3.1 Goodness-of-fit Test of Logit Model

The first step of the logistic regression is to assess the goodness of fit test of the model to the data. Statistics -2 Log Likelihood and the Hosmer-Lemeshow test usually used to assess the model fit. Let the null hypothesis is the model that fits the data and the alternative hypothesis is the model that does not fits the data. This study used Hosmer and Lemeshow test to test for the good fitting model. Based on Hosmer and Lemeshow test result in Table 1, we did not reject the null hypothesis (*p*-value = 0.146) and concluded that the model is a good fitting model at 5% significance level.

Table 1
Hosmer and Lemeshow Test

Chi-square	Df	Sig.
12.114	8	.146

3.2 Logit Model Formulation

The initial formulation of logit model consists of all the seven predictor variables. Table 2 shows the regression coefficient estimates, standard error of regression coefficient, Wald statistic, and odds ratio. The results in Table 2 indicated that only four predictors contributed significantly at 1% to the logit model. The predictors are MACD, RSI, SO and ROC. Then, the logit model is simplified by removing insignificant predictor variables and the results is shown in Table 3.

Table 2
Initial model results

	b	Standard error	Wald statistic	e^b
MACD	-37.956***	8.675	19.143	0.000
RSI	-.113***	0.043	6.929	0.893
SO	.131***	0.032	16.959	1.140
MA	.142	1.311	0.012	1.152
EMA	-1.211	1.638	0.547	0.298
ROC	.923***	0.204	20.365	2.516
VT	.025	0.174	0.021	1.025
Constant	5.044	6.559	0.591	155.153

*** Significant at 1%, ** Significant at 5%, * Significant at 10%

Table 3
Final model results

	b	Standard error	Wald	e^b
MACD	-42.295***	7.762	29.689	0.000
RSI	-0.110***	0.041	7.061	0.896
SO	0.134***	0.032	17.760	1.143
ROC	0.943***	0.197	22.810	2.568
Constant	-1.792	1.210	2.193	0.167

*** Significant at 1%, ** Significant at 5%, * Significant at 10%

Based on Table 3, all the predictors in the final logit model are significant at 1% indicating that all of them are important in predicting the category of stock movement. The model can be written as follows,

$$logit\ (p) = ln\left(\frac{p}{1-p}\right) = -1.79 - 42.30MACD - 0.11RSI + 0.13SO + 0.94ROC$$

or,

$$odds = \frac{p}{1-p} = e^{-1.79} \times e^{-42.30MACD} \times e^{-0.11RSI} \times e^{0.13SO} \times e^{0.94ROC}$$

The e^{b_i} is the odds ratio for the predictor variables and the values give the relative amount by which the odds of stock movement are going up or remain unchanged increase when the odds ratio value of SO and ROC increased by 1 unit. While, the odds of stock movement are going up or remain unchanged decrease when the odds ratio value of MACD and RSI increased by 1 unit.

3.3 In-sample Validation of Logit Model

Using the logistic regression, the goal of the analysis is to classify the observation into a particular group. Classification of observations is done by first estimating the probabilities in the indicated group, which can be computed from the logistic regression. This study used in-sample data to validate the logit model formulated from the previous section explanation. From the classification table in Table 4, it is shown that the percentage of correctly classified stock market movement is 86%.

Table 4
Classification Table

		Predicted Group		Percentage Correct
		0	1	
Observed Group	0	68	11	86.1
	1	12	73	85.9
Overall Percentage				86

3.4 Out-of-sample Validation of Logit Model

By using the different data from modelling part data, we apply the logit model into out-of-sample data for validation purpose. The results are given in Table 5.

Table 5
Estimated value for future stock market movement

Date	Actual Value	Estimated Value
01-09-2016	0	0
02-09-2016	0	1
05-09-2016	1	1
06-09-2016	0	0
07-09-2016	1	1
08-09-2016	0	1
09-09-2016	0	0

From the out-of-sample data, the study found that out of 7 days, 5 days are correctly classified. Therefore, the out-of-sample classification rate for stock market movement produces a 71.43%. This out-of-sample forecasting result is less than the in-sample prediction result that produces an 86% classification rate.

4. Conclusions

This paper successfully formulates a logit model based on the logistic regression method to predict the stock market movement. In a nutshell, by using the logistic regression method, the MACD, RSI, SO and ROC are the significant technical indicators to predict the stock market movement based on historical data from January until August 2016 has been identified. Based on the logit model, the future market movement shows a positive movement. The out-of-sample validation results show the classification rate not too far compared to in-sample validation results with classification rate value 71.43% and 86% respectively. The validation results also show the acceptable performance of logit model when more than half data are correctly classified using the formulated logit model.

Stock market plays an important role to offer unique service and benefits to businesses, individual investors, and governments. For this study, it will provide valuable information for those who are involved in the stock market. However, there is a need to determine which factors are the most influential on stock price movements to reach the target profit. This information will help to improve or maintain the performance of the price movement and get the target profits. Besides that, this

study also assists in the performance of Bursa Malaysia by providing a good model to assist the stock market in the future.

The stock market movement prediction in this paper is limited to statistical logistic regression approach only. Other approaches could be explored such as fuzzy approach [14-15], discriminant analysis [16] and multiple regression [17] to improve the forecasting accuracy with recent stock data and more significant technical indicators.

Acknowledgment

The authors are grateful for the financial support received from Universiti Utara Malaysia under the Research Generating University Grant (S/O Code: 13878). We also thank the anonymous reviewers for their useful suggestions.

References

[1] Weckman, Gary R., and Ranjeet Agarwala. "Identifying relative contribution of selected technical indicators in stock market prediction." In *IIE Annual Conference. Proceedings*, p. 1. Institute of Industrial and Systems Engineers (IISE), 2003.

[2] Preethi, G., and B. Santhi. "STOCK MARKET FORECASTING TECHNIQUES: A SURVEY." *Journal of Theoretical & Applied Information Technology* 46, no. 1 (2012).

[3] Qin, Qin, Qing-Guo Wang, Shuzhi Sam Ge, and Ganesh Ramakrishnan. "Chinese stock price and volatility predictions with multiple technical indicators." *Soft-Computing in Capital Market: Research and Methods of Computational Finance for Measuring Risk of Financial Instruments* (2014): 109.

[4] Paliyawan, P. "Stock Market Direction Prediction Using Data Mining Classification." *ARPN Journal of Engineering and Applied Sciences* 10, no. 3 (2015): 1302-1310.

[5] Al-Radaideh, Qasem A., Adel Abu Assaf, and Eman Alnagi. "Predicting stock prices using data mining techniques." In *The International Arab Conference on Information Technology (ACIT'2013)*. 2013.

[6] Olaniyi, S. Abdulsalam Sulaiman, Kayode S. Adewole, and R. G. Jimoh. "Stock trend prediction using regression analysis–a data mining approach." *ARPN Journal of Systems and Software* 1, no. 4 (2011): 154-157.

[7] Agrawal, J. G., V. S. Chourasia, and A. K. Mittra. "State-of-the-art in stock prediction techniques." *International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering* 2, no. 4 (2013): 1360-1366.

[8] Dutta, Avijan, Gautam Bandopadhyay, and Suchismita Sengupta. "Prediction of stock performance in indian stock market using logistic regression." *International Journal of Business and Information* 7, no. 1 (2012).

[9] Samant, Shruti. "Prediction of Financial Performance Using Genetic Algorithm and Associative Rule Mining." *International Journal of Engineering Research and General Science* 3, no. 1 (2015): 1035-1045.

[10] Fayek, Magda B., Hatem M. El-Boghdadi, and Sherin M. Omran. "Multi-objective optimization of technical stock market indicators using gas." *International Journal of Computer Applications* 68, no. 20 (2013).

[11] Gamil, Ahmed Abdullah, Raafat S. Elfouly, and Nevin M. Darwish. "Stock Technical Analysis using Multi Agent and Fuzzy Logic." In *World Congress on Engineering*, no. 4, p. 6. 2007.

[12] Preethi, G., and B. Santhi. "STOCK MARKET FORECASTING TECHNIQUES: A SURVEY." *Journal of Theoretical & Applied Information Technology* 46, no. 1 (2012).

[13] Caley, Jeffrey Allan. "A Survey of Systems for Predicting Stock Market Movements, Combining Market Indicators and Machine Learning Classifiers." (2013).

[14] Mansor, Rosnalini, Mahmod Othman, and Maznah Mat Kasim. "Fuzzy Reasoning Based Time Series Forecasting Using Weighted Subsethood Model." *Advanced Science Letters* 23, no. 9 (2017): 9094-9097.

[15] Mansor, Rosnalini, Maznah Mat Kasim, and Mahmod Othman. "The framework of weighted subset-hood Mamdani fuzzy rule based system rule extraction (MFRBS-WSBA) for forecasting electricity load demand." In *AIP Conference Proceedings*, vol. 1761, no. 1, p. 020061. AIP Publishing, 2016.

[16] Alrasheedi, Melfi, and A. Alghamdi. "Predicting up/down direction using linear discriminant analysis and logit model: the case of SABIC price index." *Research Journal of Business Management* 6, no. 4 (2012): 121-133.

[17] Smith, Timothy A., and Alcuin Rajan. "A Regression Model to Predict Stock Market Mega Movements and/or Volatility Using Both Macroeconomic Indicators & Fed Bank Variables." *International Journal of Mathematics Trends and Technology*49, no. 3 (2017): 165.