Multiple Linear Regression for Technical Outlook in Telecom Stock Price

Pichit Boonkrong*, Nithipa Arjrith and Seksan Sangsawad

College of Digital Innovation and Information Technology,
Rangsit University, Pathum Thani, 12000, THAILAND
*Corresponding author, E-mail: pichit.bk@rsu.ac.th

Abstract

In this paper, a multiple linear regression is employed as a technical tool to analyze the relationship among gold price, Dow Jones industrial average, oil price and the trading values in Advanced Info Service Public Company Limited (AIS) during January 2- July 31, 2019 (141 days). Both descriptive and inferential statistics are accordingly presented. In order to implement the multiple linear regression model, suitability, essential assumptions and hypotheses are statistically examined. At a significance level of 0.05, it is found that only gold price and Dow Jones industrial average have the linear relationship with the trading values in AIS Company with the coefficient of determination of 0.998. Therefore, the guideline for trading on the stock of AIS Company Limited can be obtained from the signals of gold price and Down Jones industrial average.

Keywords: *Dow Jones, Model fitting, Multiple regression, Stock market, Technical analysis, Telecommunication.*

1. Introduction

Over the technological advance of the 21st century, telecommunication becomes more and more important as the demand of transmitting information by electromagnetic means is dramatically increasing. Thus, the investment in modern digital telecommunication is also growing. However, it is still questioned that which factor has an impact on the stock price of the telecommunication companies.

Analyzing influencing factors, multiple linear regression (MLR) is presented as an effective tool for solving problems related to various environmental factors. To forecast appropriate points for various activities, such as exploring for natural resources (Hernández et al, 2016; Peng et al, 2016) or maintenance work in large industries (Abbasi et al, 2018), In business, the variation in the factor of interest can be captured by explanatory variables. MLR is also offered as a basis for trading securities (Sharma et al, 2017). Izzah developed applications predicting stock prices in real time using improved multiple linear regression (IMLR) and moving average (MA) as analysis tools (Izzah et al, 2017). Akşehir & Kılıç used MLR to predict the price of stocks in the bank sector using 50 indicators divided into 46 technical and 4 fundamental indicators (Akşehir & Kılıç, 2019). It is advantageous that MLR could appropriately reduce the number of indicators and it remained only 29 important technical factors. Regarding to stock trading, MLR is also in a part of many researches for assessment, prediction and evaluation (Jia et al, 2019; Li et al, 2015; Nivetha & Dhaya, 2017). Technically, there are some interesting factors including securities index such as NASDAQ, Russell 2000, Dow Jones or S&P500 and oil prices which are often added into a forecasting model of stock price (Bekiroglu et al, 2018; Eapen et al, 2019). Dow Jones industrial average is calculated from the top 30 influential companies in the USA and it is considered as the representative of the US stock. Thus, Dow Jones is the key indicator for global stock markets. Likewise, the business growth directly varies to the oil price, i.e. whenever there is the growth in business values, the demand in oil consumption also increases. It follows that the oil price is one of the key factors behind any stock variation. Additionally, the gold price is also included in a stock assessment as gold is intuitively considered as the most secure asset. People usually buy gold when there is any crisis, especially financial crisis. Considering the stock of the telecom sector in Thailand, most of them are always ranked into SET50 index (Arjrith & Boonkrong, 2019). Therefore, the main aim of this study is to formulate the MLR model to investigate the influence of Dow Jones index, oil price and gold price on the stock price of the telecom sector, particularly in Advanced Info Service Public Company Limited (AIS).

The implementation of MLR model, a very powerful statistical technique, has been widely used to analyze the relationship between response and explanatory variables. It is the main research tool in this paper. The presentation of this paper is organized as follows. The MLR model and its assumptions are presented in Section 2. The empirical results from the given dataset and MLR model are given in Section 3, followed by discussion and conclusion in Section 4.

2. Research Methodology

This section illustrates the implementation of a statistical model for studying the relation between response and explanatory variables as shown in Table 1.

Table 1 Descriptions of variables

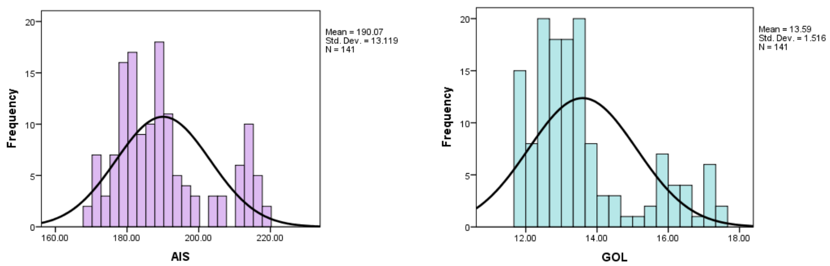
Variables	Descriptions	Unit	Duration
AIS	Trading value of AIS Co., Ltd.	Million Baht	January 2 – July 31, 2019 (Sample size = 141 days)
OIL	Stock's closing price of oil	Million USD	
GOL	Stock's closing price of gold	Million USD	
DJS	Dow Jones industrial average	-	

2.1 Data Source

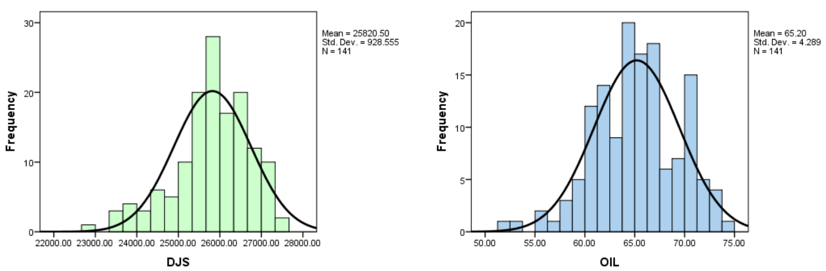
The data used to formulate the multiple linear model includes AIS, Oil, Gold and Dow Jones during January 2- July 31, 2019 (141 days). The descriptive statistics including mean, standard deviation, minimum and maximum of each variable are given in Table 2 and their histograms are shown in Figure 1.

Table 2 Descriptive statistics of variables

Variables	n	Minimum	Maximum	Mean	Std. Dev
AIS	141	169.00	218.00	190.07	13.12
OIL	141	52.14	74.14	65.20	4.29
GOL	141	11.74	17.35	13.59	1.52
DJS	141	22686.22	27359.16	25820.50	952.55



(a) (b)



(c) (d)

Figure 1 Histograms of variables used in multiple linear regression analysis

2.2 Multiple Linear Regression

As three explanatory variables are added into consideration, a multiple linear regression model is used. It is firstly assumed that the trading value in AIS is influenced by oil, gold and Dow Jones. Then, it yields

$$\text{AIS} = \beta_0 + \beta_1\text{GOL} + \beta_2\text{DJS} + \beta_3\text{OIL} + \varepsilon \quad (1)$$

where β_0, β_1, β_2 and β_3 respectively denote the intercept and parameters for each explanatory variable and $\varepsilon \sim N(0,1)$ denotes the error. To apply MLR analysis, the following four assumptions must be satisfied.

- The explanatory and response variables have linear relationship to each other.
- The explanatory variables are not highly correlated with each other (no multicollinearity).
- The residuals are normally distributed (multivariate normality).
- The variance of error terms are similar (homoscedasticity).

To be more specific, the effect of explanatory variables are considered through the parameters β_0, β_1, β_2 and β_3. The estimation of parameters in model (1) is carried out using MLR analysis and the more details are illustrated in the next section.

3. Empirical Results

To apply stepwise multiple regression as presented in Section 2, an initial screening of the candidate variables is carried out. Firstly, it needs to examine whether there is a linear relationship between explanatory and response variables. The correlation between each explanatory variable and response variable is examined by the following three hypotheses.

Hypothesis 1

H_0 : GOL has no relationship with AIS.

H_1 : GOL has relationship with AIS.

Hypothesis 2

H_0 : DJS has no relationship with AIS.

H_1 : DJS has relationship with AIS.

Hypothesis 3

H_0 : OIL has no relationship with AIS.

H_1 : OIL has relationship with AIS.

Table 3 Pearson correlations between variables GOL, DJS, OIL and AIS

Variables	GOL	DJS	OIL
AIS	0.851	0.796	0.190
Sig. (2-tailed)	0.000*	0.000*	0.024*
N	141	141	141

* $p < 0.05$

The relationship between variables are examined by Pearson correlations as given in Table 3. It is found that all aforementioned hypotheses reject H_0. That is, GOL, DJS and OIL have relationships with AIS at the significance level of 0.05. Next, it is the time to perform multiple linear regression analysis. Using the stepwise multiple linear regression method, OIL is removed from the model (1) as its tolerance (0.290) and variance inflation factor (3.451) are inappropriate. Thus, it yields

$$\text{AIS} = \beta_0 + \beta_1\text{GOL} + \beta_2\text{DJS} + \varepsilon. \quad (2)$$

Then, the impact of GOL and DJS on AIS are examined via the following hypothesis.

Hypothesis 4

$H_0: \beta_1 = \beta_2 = 0$

$H_1: \beta_1 \neq 0$, or $\beta_2 \neq 0$

Considering ANOVA in Table 4, it is found that P-value = 0.00 < 0.05, i.e., the null hypothesis H_0 is rejected and the alternative hypothesis H_1 is accepted. It is noted that both parameters β_1 and β_2 are not zeros.

Table 4 ANOVA

Model	Sum of Squares	df	Mean Square	F	Sig.
Regression	19551.190	2	9775.594	296.855	0.000*
Residual	4544.421	138	32.931		
Total	24095.610	140			

* p < 0.05

Table 5 Estimation of Parameters

Model	Unstandardized Parameters		Standardize Parameters	t	Sig.	Collinearity Statistics	
	B	Std. Error	Beta			Tolerance	VIF
Constant	-24.952	15.013		-1.662	0.099		
GOL	4.988	0.436	0.576	11.431	0.000*	0.538	1.860
DJS	0.006	0.001	0.404	8.003	0.000*	0.538	1.860

* p < 0.05

Since the value of tolerance is 0.538 > 0.2 and the variance inflection factor is 1.860 < 5 as given in Table 5, it is verified that there is low multicollinearity among variables. Next step, it is important to consider if each parameter should be included to the model via the following hypotheses:

Hypothesis 5

$H_0: \beta_0 = 0$

$H_1: \beta_0 \neq 0$.

Hypothesis 6

$H_0: \beta_1 = 0$

$H_1: \beta_1 \neq 0$.

Hypothesis 7

$H_0: \beta_2 = 0$

$H_1: \beta_2 \neq 0$.

Regarding to the results in Table 5, the null hypothesis is rejected for Hypothesis 6 and Hypothesis 7 as P-value = 0.000 < 0.050. Therefore, both parameters β_1 and β_2 are still kept in the model whereas the constant β_0 is removed as P-value = 0.099 > 0.050 in Hypothesis 5. Subsequently, the model is rewritten as follows:

$$AIS = \beta_1 GOL + \beta_2 DJS + \varepsilon \quad (3)$$

Table 6 Estimation of Parameters without constant

Model	Unstandardized Parameters		Standardize Parameters	t	Sig.
	B	Std. Error	Beta		
GOL	5.306	0.000	0.620	21.864	0.000*
DJS	0.005	0.395	0.381	13.441	0.000*

* $p < 0.05$

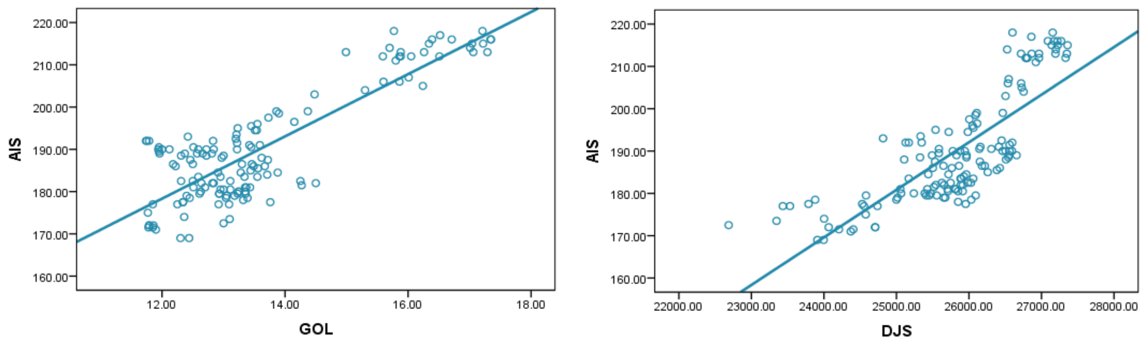
Performing regression without constant in SPSS software, the values of parameters β_1 and β_2 are figured out as shown in Table 6. The model (3) is eventually specified by

$$\hat{AIS} = (5.306)GOL + (0.005)DJS \quad (4)$$

where $\hat{AIS}$ denotes the approximation of AIS with unstandardized parameters. Alternatively, it can also use the standardized parameters to express such linear relationship, i.e.

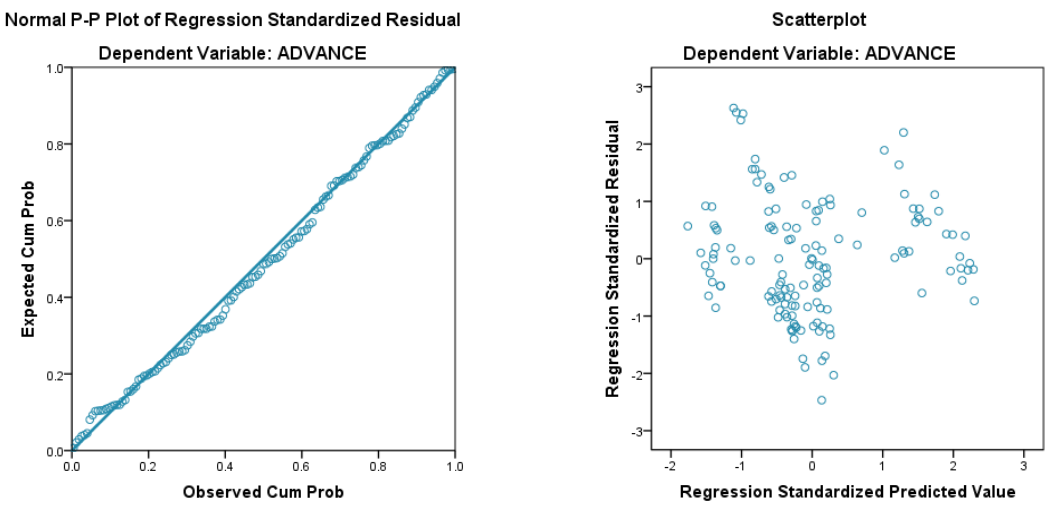
$$Z_{AIS} = (0.620)Z_{GOL} + (0.381)Z_{DJS}. \quad (5)$$

In model summary, the measure of goodness of fit for model (4) is based on the coefficient of determination $R^2 = 0.998$, i.e., GOL and DJS have a high level of linear relationship with AIS. To accomplish the assumption of the multiple linear regression model, the linear relationship between response variable and explanatory variables is investigated by scatterplots. The linear relationship between GOL and AIS is presented in Figure 2(a). The linear relationship between DJS and AIS is presented in Figure 2(b). The change in GOL value for 1 unit can directly vary AIS value for 5.306 units and the change in DJS for 1 unit can directly vary AIS value for 0.005 units.



(a) (b)

Figure 2 Scatterplots presenting the linear relationships between response and explanatory variables



(a) (b)

Figure 3 Normal P-P plot of regression standardized residual and predicted value

As can be seen in Figure 3(a), the residuals are normally distributed. The plot of standardized residuals versus predicted values in Figure 3(b) shows that the points are equally distributed across all values of the explanatory variables, i.e. the range of standardized residuals lies between -3 and 3. It means that the homoscedasticity of model (3) is claimed. Therefore, all four assumptions as mentioned at the end of Section 2 are satisfied for MLR analysis. To see the difference between the actual values of AIS and model (4), their line graphs are plotted and displayed in Figure 4.

4. Discussion and Conclusion

This paper has introduced a novel multiple linear regression model to analyze the relationship among gold price, Dow Jones industrial average, oil price and the trading value in AIS Company. To follow the multiple linear regression model, four assumptions including linear relationship, multicollinearity, multivariate normality and homoscedasticity are verified. At the significance level of 0.05, six hypotheses corresponding to MLR analysis and the dataset are examined. Regarding hypotheses 1-3, all explanatory variables are correlated to the response variable, i.e. GOL, DJS and OIL are correlated to the value of AIS. By hypotheses 4-7, it is guaranteed that only GOL and DJS have the effect on the value of AIS as their parameters β_1 and β_2 are nonzero where the constant β_0 is removed from the model. Using the MLR model and its assumption, the linear equation describing the relationship between response and explanatory variables is finally obtained as $\hat{AIS} = (5.306)GOL + (0.005)DJS$ with the coefficient of determination $R^2 = 0.998$. The technical outlooks including statistical and graphical evidences are accordingly presented. The contribution of this paper can support investors in making decisions when trading in telecommunication stocks, i.e. the signals from gold price and Dow Jones should be taken into account as they have a strong linear relationship to each other. However, trading in stock market is usually risky. If there is an economic crisis, people will feel more secure to invest on the actual gold because it is intuitively considered as a major economic driver in every country.

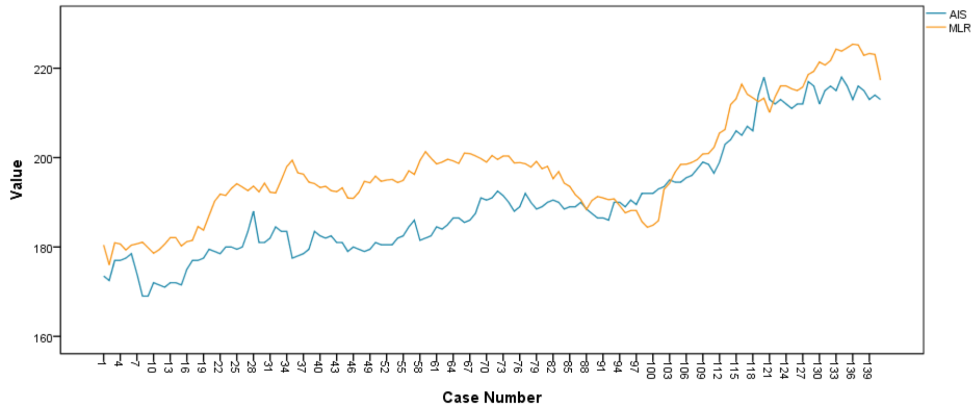


Figure 4 Comparison between the actual values of AIS and multiple linear regression (MLR) model.

6. References

Abbasi, T., Lim, K. H., Rosli, N. S., Ismail, I., & Ibrahim, R. (2018). Development of Predictive Maintenance Interface Using Multiple Linear Regression. 2018 International Conference on Intelligent and Advanced System, 1–5. https://doi.org/10.1109/ICIAS.2018.8540602

Akşehir, Z. D., & Kılıç, E. (2019). Prediction of Bank Stocks Price with Reduced Technical Indicators. *2019 4th International Conference on Computer Science and Engineering*, 206–210. https://doi.org/10.1109/UBMK.2019.8906999

Arjrith N., & Boonkrong, P. (2019). Analysis of Influencing Factors for Trading Value in Telecommunication Company, *Proceedings of RSU Research Conference*, 434–439.

Azam, M., & Kumar, D. (2011). Factors Influencing the Individual Investor and Stock Price Variation: Evidence from Karachi Stock Exchange, *Australian Journal of Basic and Applied Sciences*, 5(12): 3040–3043.

Bekiroglu, K., Duru, O., Srinivasan, S., Su, R., Lagoa, C., & Li, J. (2018). Governing Dynamics of Crude Oil and LNG Prices. *IEEE 14th International Conference on Control and Automation*, 872–877. https://doi.org/10.1109/ICCA.2018.8444188

Dobson, Annette J. (1945). An introduction to generalized linear models. Boca Raton : Chapman & Hall/CRC.

Eapen, J., Bein, D., & Verma, A. (2019). Novel Deep Learning Model with CNN and Bi-Directional LSTM for Improved Stock Market Index Prediction. *IEEE 9th Annual Computing and Communication Workshop and Conference*, 264–270. https://doi.org/10.1109/CCWC.2019.8666592

Hernández, E., Vera, E., & Vasquez, P. (2016). Mid-term load forecasting - application in oil facilities. *IEEE Industry Applications Society Annual Meeting*, 1–6. https://doi.org/10.1109/IAS.2016.7731821

Izzah, A., Sari, Y. A., Widyastuti, R., & Cinderatama, T. A. (2017). Mobile app for stock prediction using Improved Multiple Linear Regression. *International Conference on Sustainable Information Engineering and Technology*, 150–154. https://doi.org/10.1109/SIET.2017.8304126

Jia, L., Wang, F., Sun, H., & Li, H. (2019). An Empirical Research of the Impact of Monetary Policy on Stock Market Returns—Based on the perspective of investor structure. *Chinese Control and Decision Conference*, 4158–4163. https://doi.org/10.1109/CCDC.2019.8833475

Li, C., Miao, Y., Sun, R., & Sun, Z. (2015). The study of the refined oil price trend based on AHP and GM method. *The 27th Chinese Control and Decision Conference*, 4430–4434. https://doi.org/10.1109/CCDC.2015.7162654

Nivetha, R. Y., & Dhaya, C. (2017). Developing a Prediction Model for Stock Analysis. *International Conference on Technical Advancements in Computers and Communications*, 1–3. https://doi.org/10.1109/ICTACC.2017.11

Peng, Z., Yang, H., Pan, H., & Ji, Y. (2016). Identification of low resistivity oil and gas reservoirs with multiple linear regression model. *The 12th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery*, 529–533. https://doi.org/10.1109/FSKD.2016.7603229

Sharma, A., Bhuriya, D., & Singh, U. (2017). Survey of stock market prediction using machine learning approach. *2017 International Conference of Electronics, Communication and Aerospace Technology*, 2: 506–509. https://doi.org/10.1109/ICECA.2017.8212715