

The logit and fitting analysis of logistic regression on jakarta composite index in covid-19 pandemic

Mutijah

IAIN Purwokerto, Jl. Jend. A. Yani, No.40A, Purwokerto, 53126

mutijah1972@gmail.com

Abstract. Many statistic study about forecasting, predictive, estimation, or relationship have been developed globally either before or during the covid-19, including in Indonesia. This study purposes to analyze the odds and fitting of the relationship between the increasing and decreasing daily of covid-19 cases with the up and down of Jakarta Composite Index. Logistic regression method was used to analyze them. Logistic regression, also called logit regression or logic model was one of the methods to analyze in determining of the relationship between two or more variables which the dependent variable is category and the independent variables can be categorical, continuous, or a combination. The relationship can be in variables between economy and physics, economy and agriculture or biology and medical, etc. The odds played an important role in logistic regression. In this study, the odds was a good indicator that showed the opportunity of up against down under the data of Jakarta Composite Index. The analysis were done on variable of the increasing and decreasing daily of covid-19 cases in continuous and categorical. This analysis were based on Jakarta Composite Index historical data for 120 days on March 17, 2020 until September 16, 2020. The analysis results were obtained that the odds established an up was 1,267 times as likely as a down in continuous independent variable and 1,264 times in categorical. While the probability established the logit model was not fit with the values of goodness-of-fit very low. This matter did not mean that the logic model was made to produce a false regression. The value of goodness-of-fit on variable of the increasing and decreasing daily of covid-19 cases in continuous is greater than in categorical.

1. Introduction

The global covid-19 pandemic has affected many sectors, including economy, education, agriculture, health and others. The covid-19 pandemic has not yet been determined when it will end, although several research of the forecasting, predictive, and estimation have been carried out to be this importance by mathematicians and statisticians in the world, including in Indonesia. Based on the opinion that the covid-19 pandemic was affecting in many sectors, a relationship research can also be done. Although similar research had also been done before the Covid-19 pandemic. Especially during the covid-19 pandemic, there may be the interesting variables to be researched in its relationship. For example, the relationship between the increasing and decreasing of covid-19 daily cases with the stock price index in a country by using binary logistic regression. Logistic regression is mathematical modelling approach used to define the relationship between independent variables and dependent variables. The dependent variable is established in categories. Further, the dependent variable which it was in two categories so that was namely binary logistic regression. Several research about the relationship and the affecting by using logistic regression before covid-19 pandemic have been done the world researcher or Indonesian.

Pyke and Sheridan [1] investigated the retention of master and doctoral candidates was whether the student successfully completed the degree affected by selected demographic, academic, and financial support variables. Korkmaz, Guney, and Yigiter [5], their study resulted the applicability of logistic regression in the Turkish livestock sector. Ramosacaj, Hasani, and Dumi [7] researched that the examination of student's performance from university of Vlore was affected by gender. Tukur and Usman [9] applied the binary logistic regression analysis technique to form a fitting model for data and significance of the two independent variables that was gender and Jamb scores. Zewude and Ashine [11] analysed in assessment and identifying factors that influence student's academic achievement by using binary logistic regression. Study from Francis [12] investigated and modelled salient factors that influences the performance of staff in an appraisal exercise and as well the odds of these factors influencing the outcome variable as compared to their reference category. Ajao and Sanni [16] examined the relationship between selected demographic and socioeconomic variables and current use of contraception using logistic regression technique. Isa, Ikughur, Mohammed, and Ovyne [19] assess the quality of life of civil servant in relation to socioeconomic in Nasarawa State by using binary logistic regression model. Astari and Kismiantini [17] determined factors affecting the health assurance ownership and to understand their relation in Indonesia by using binary logistic regression model. Nurdiansah and Khikmah [23] analyzed variables that influence poverty in Central Java, Indonesia with binary logistic regression. While research was not a relationship that is forecasting, predicting, and estimation have been done the world researchers or Indonesia before or during the covid-19 pandemic using logistic regression technique or others. Subagyo and Sugiarto [8] use the logit model to estimated the using of different commands to the data analysis company with CSR performance ratio as a variable. Zaidi and Amirat [10] forecasted the stock market trends by using logistic regression model and Neural Networks. Chattopadhyay Majumdar and Prasad [15] predicted such channel decisions based on binary logistic regression modelling approach. Abonazel and Ibrahim [14] reviewed estimation methods for binary logistic regression model with missing values. The results indicated that the regression imputation method was very appropriate. Lih and Ismail [20] applied binary logistic regression to predict the influence of teacher efficacy on literacy and numeracy. Durica, Valaskova and Janoskova [18] predicted the business failure of companies operating in V4 Countries based on logistic regression analysis. Verma [21] predicted the probability of corporate default using logistic regression. The two of research during the covid-19 pandemic that have been done in Indonesia were by Nuraini, Khairudin and Apri [22] proposing a simple model to predict the endemic in Indonesia and Zuhairoh and Rosadi [24] forecasted the real-time of the covid-19 epidemic using the Richards Model in South Sulawesi, Indonesia.

The IDX Composite or Jakarta Composite Index (JSX) is an index of all stocks that are traded on the Indonesia Stocks Exchange (BEI). This index was an indicator of stock price movements on the stock market of the Indonesia Stock Exchange (BEI). This index covered the price movements of all common shares and preferred shares listed on the Indonesia Stock Exchange. The base day for Jakarta Composite Index is August 10, 1982. The Indonesia Stock Exchange set the index with a base value of 100 and records 13 shares on that date. The basis for calculating Jakarta Composite Index was the total market value of shares listed on August 10, 1982. The index calculation represents the movement of shares that occur in the market. The occurrence of a global covid-19 pandemic affected the movement of stock prices on world stock markets. Likewise the covid-19 pandemic affected the movement of share prices on the Indonesia Stock Exchange, which can be seen from the up and down of Jakarta Composite Index. Thus we could think about depending on the Jakarta Composite Index, whether up or down with the increasing or decreasing of covid-19 cases. We could also determine the number of times that events are more likely to go up or down. In this phenomenon we could describe the event by analyzing of the odds on logistic regression. We would also analyze whether logistic regression is fit or not to describe the phenomenon by the R software application. Related to this study, in Indonesia has been researched in Mutijah [13] which modelled the stock price index in a mixed method that was compared with a simple aggregative price index on binary logistic regression model.

2. Methods

2.1. Data source

Data for this research work were the data of Jakarta Composite Index and the increasing and decreasing daily of covid-19 on 120 days in Indonesia. They were taken on March 17, 2020 until September 16, 2020. The data of Jakarta Composite Index were obtained from Indonesia Stock Exchange (BEI) in www.yahoo.finance. The data of the increasing and decreasing daily of covid-19 in Indonesia were obtained from www.google.com by giving of key words were corona case in Indonesia on google site.

2.2. Variable used

The dependent variabel, namely Y in this research was the up and down of Jakarta Composite Index which is a category. The dependent variable category 1 signed the up and 0 was the down in Jakarta Composite Index. The independent variable, namely X in this research was the increasing and decreasing daily of covid-19 in Indonesia. This independent variable was analysed in continuous and category.

2.3. The logit and fitting analysis of logistic regression

Logistic regression is a statistical modeling technique which the probability of categorical dependent variable is related to the independent variables in continuous or category. In general, the logistic regression probability function is presented as form

$$\pi(x) = \frac{\exp(\beta_0 + \beta_1 x)}{1 + \exp(\beta_0 + \beta_1 x)} \quad (1)$$

$\pi(x)$ is the probability of success category on the dependent variable and is estimating parameter for the independent variable. Further, the corresponding logistic regression model form is

$$\log\left(\frac{\pi(x)}{1 - \pi(x)}\right) = \alpha + \beta x \quad (2)$$

The logistic regression model is a special case of a Generalized Linier Model (GLM) which the random component for the (success, failure) outcomes has a binomial distribution. The link function is the logit function $\log(\pi/(1-\pi))$ of π and was symbolized by $\log(\pi)$. For probability of success π , as it is known that the odds of success are defined to be

$$\pi = \frac{\pi}{1 - \pi} \quad (3)$$

The odds are nonnegative that is a success is more likely than a failure. It can be seen that the success probability is the function of odds, that is

$$\pi = \frac{\text{odds}}{\text{odds} + 1} \quad (4)$$

3. Results

3.1. The logit and fitting analysis of logistic regression on the continuous dependent variable.

The logit and fitting analysis was used to determine the up probability of Jakarta Composite Index through its odds and the fitting of logistic regression. In this matter, it was analysed on the continuous dependent variable. It was carried-out by using the R Command Line Interface (RCLI) package including summary of odds ratio, summary of logit regression, and value of goodness of fit. The odds ratio analysis was used to know how many times the Jakarta Composite Index was the up as likely as the down. Tabel 1. referred a results of odds ratio summary.

Tabel 1. Analysis results of odds on R command line interface (RCLI)

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
1.171	1.196	1.251	1.267	1.313	1.497

While analysis of the fitting of logistic regression was used to know the significant of probability. Tabel 2. presents the results of logit regression summary.

Tabel 2. Analysis results of logit regression on R command line interface (RCLI)

<u>Coefficients:</u>				
	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	1.556e-01	2.944e-01	0.528	0.597
Covid-19	6.262e-05	1.833e-04	0.342	0.733

Analysis later, it determined the goodness of fit. This analysis was carried out to see the fit of the regression model form. The value of goodness of fit was obtained 0.0007108123. The goodness of fit is too small to indicate a model fit.

3.2. The logit and fitting analysis of logistic regression on the category dependent variable.

Analogously on subsection 3.1, this analysis was also used to determine the up probability of Jakarta Composite Index through its odds and the fitting of logistic regression. But in this matter, it was analysed on the category dependent variable. It was also carried out by using the R Command Line Interface (RCLI) package including summary of odds ratio, summary of logit regression, and value of goodness of fit. Tabel 3. referred a results of odds ratio summary.

Tabel 3. Analysis results of odds on R command line interface (RCLI)

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
1.250	1.250	1.250	1.264	1.286	1.286

Analysis of the fitting of logistic regression was also worked to know the significant of probability in this subsection. Tabel 4. also presents the results of logit regression summary.

Tabel 4. Analysis results of logit regression on R command line interface (RCLI)

<u>Coefficients:</u>				
	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	0.25131	0.29095	0.864	0.388
Covid-19	-0.02817	0.37537	-0.075	0.940

The same as in subsection 3.1 was also analysed to determine the goodness of fit. This analysis was also carried out to see the fit of the regression model form. The value of goodness of fit was obtained $3.420085e-05$. The goodness of fit is also too small to indicate a model fit.

4. Discussion

The study as in above give that there were the matter to be discussed that is this research attempt to analyze the odds and the up probability on the logistic regression or it is called logit model describing the relationship between the increasing and decreasing daily covid-19 case with Jakarta Composite Index in Indonesia for the continuous independent variable and category. This research also attempt to analyze the fitting of its logistic regression. The odds and the up probability for the continuous independent variable is greater than the category. The fitting probability is not significant and the value of goodness of fit is very low which the value of goodness of fit for the continuous independent variable is greater than the category. The other study also showed the fitting probability is not significant and the value of goodness of fit is very low for the continuous independent variable is greater than the category, but it does not mean that a regression was made to produce a false logistic regression. This corresponded finding has resulted the same of results in Subagyo and Sugiarto [8] that is the probability of logit model is not fit and the value of adjusted R squared very low.

5. Conclusion

Based on the analysis results of the odds, the up probability, and the fitting of logistic regression or it is called logit model can be concluded that the odds and the up probability of logistic regression on the continuous independent variable is greater than on the category, the significant of probability is greater than 0,05 and the value of goodness of fit is very low so that is resulted the logistic regression is not fit and the fitting of logistic regression on the continuous independent variable is better than on the category.

References

- [1] Pyke SW & Sheridan PM 1993 *The Canadian Journal of Higher Education*. Vol. XXIII-2
- [2] Hosmer D W and Lemeshow S 2000 *Applied Logistic Regression Second Edition* (New York: John Wiley & Sons.Inc)
- [3] Agresti A 2007 *An Introduction to Categorical Data Analysis Second Edition* (Hoboken New Jersey:John Wiley & Sons.Inc)
- [4] Tranmer M and Elliot M 2008 *Binary Logistic Regression* (Cathie Marsh Centre for Census and Survey Research) pp 6-8
- [5] Korkmaz M, Guney S, Y and Yigiter S Y 2012 *J.Agric.Fac.HR.U.* 2012, 16(2):25-36
- [6] Branimir M, Radivoje P and Predrag D 2015 *Binary Logistic Regression Modeling of Idle CO Emissions in Order to Estimate Predictors Influences in Old Vehicle Park* (Hindawi Publishing Corporation) 10 pages
- [7] Ramosacaj M, Hasani V and Dumi A 2015 *Journal of Educational and Social Research*. Vol.5 No. 3
- [8] Subagyo A and Sugiarto T 2016 *International Journal of Advanced and Applied Sciences*. 3(1) 2016, pages: 15-21
- [9] Tukur K and Usman A U 2016 *International Journal of Current Research* Vol. 8, Issue, 01, pp.25235-25239
- [10] Zaidi M and Amirat A 2016 *Euro-Asian Journal of Economics and Finance* Volume 4, Issue 2, pages: 50-58
- [11] Zewude B T and Ashine K M 2016 *Journal of Education and Practice*. Vol. 7, No. 25
- [12] Francis R M 2017 *Science Journal of Applied Mathematics and Statistics*.2017:5(4):164-168
- [13] Mutijah 2017 *Proc. Int. Conf. on Mathematics and Mathematics Education* (Universitas Ahmad

- Dahlan, Yogyakarta) p 82-87
- [14] Abonazel M R and Ibrahim M G 2018 *International Journal of Mathematics and Computational Science*. Vol. 4, No. 3, pp 79-85
 - [15] Chattopadhyay A, Majumdar M, and Prasad G V S 2018 *IOSR Journal of Business and Management (IQSR-JBM)*.pp 54-58
 - [16] Ajao M O and Sanni M O O 2019 *International Journal of Statistics and Applications*. 2019,9(5):129-133
 - [17] Astari D W and Kismiantini 2019 *IOP Conf. Series:Journal of Physics: Conf. Series* 1320(2019) 012011
 - [18] Durica M, Valaskova K, and Janoskova K 2019 *International Society and Manufacturing Service and Management Engineering*. Volume 11, Issue 4, pages: 54-64
 - [19] Isa E M, Ikughur J A, Mohammed R Y and Ovy A 2019 *International Journal of Applied Science and Mathematical Theory*. Vol. 5, No. 2
 - [20] Lih J S J and Ismail R 2019 *World Journal of Education*. Vol. 9, No. 1
 - [21] Verma D 2019 *UGC Approval No:40934 Vol.3, Issue-1, Addendum 7 (Special Issue)*.pages 161-173
 - [22] Nuraini N, Khairudin K, and Apri M 2020 *COMMUN. BIOMATH. SCI*. VOL.3, NO. 1, PP. 1-8
 - [23] Nurdiansyah S N and Khikmah L 2020 *J Int Comp & He Inf*. Vol. 1, No. 1
 - [24] Zuhairoh F and Rosadi D 2020 *Indonesian Journal of Science and Technology* 5(3), 456-462