Comparison of the binary logistic regression model on the stock price index

Mutijah

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

E-mail : mutijah1972@gmail.com

Abstract. The stock price index is the stock price stated in the index number. There are several methods for determining the price index, including the unweighted/simple aggregative and the chain price index methods. This paper proposes a mixed method between the two methods to determine the stock price index. The mixed method calculate price index based on the simple aggregative method and determining up or down for stock price based on the chain price index method. This paper also attempts to model the stock price index by a mixed method into a binary logistic regression model and compares a binary logistic regression model formed by the simple aggregative price index method. The result of research for Microsoft stock data since 10 years shows that a binary logistic regression model from the stock price by a mixed method is better than a binary logistic regression model by a simple aggregative method.

1. Introduction

The index number is a statistical measure showing the change of a variable or a set of variables related to each other, either at the same time or place. The index number is a relative number expressed as a percentage. There are three methods of calculating the index numbers: weighted/simple index numbers consisting of simple and relative aggregative methods, weighted index numbers, and chain index numbers. Simple index number method includes method of calculating price, quantity, and value index number. While the method of weighted index numbers includes the method of Aggregation, Laspreyres, Paasche, Fisher, Drobish and Bowley, Marshall-Edgeworth, and Walsh.

The price index is a number that indicates a change in the price, either a single or various commodity, in the same time or place. One of the things that need to be attented in the calculation of the price index number is the selection of the base period. The base period or so-called base year is a period or year whose index number is 100 or 100%, while the next year as a certain year. The calculation of the price index can be done on the price with one commodity or several commodities. The calculation of the price index with one commodity can be done by the selection of a fixed or non-fixed base year. Furthermore, the calculation of the price index with one commodity with a non-fixed base year is called the chain index number method.

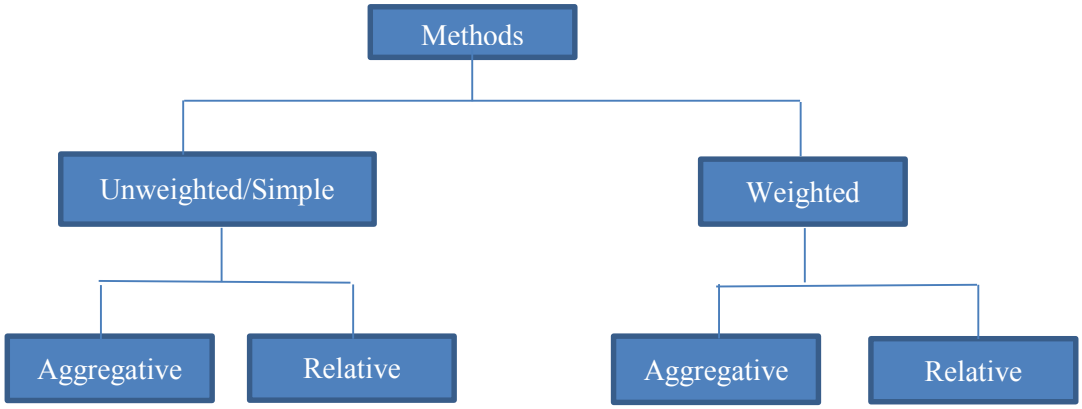
A particular stock can be interpreted as one particular commodity. Therefore the stock price index can be calculated by using simple aggregative and chain index number methods. The calculation of simple aggregative index number is calculated from the amount of the price in a given year divided by the amount of the price in the base year specified. While the calculation of chain index number is calculated from the price in a given year divided by the price in the previous year. In this case the stock price index can be done by calculating the amount of stock price in a given year divided by the

amount of stock price in the previous year. In other words, calculation can be done from a combination of simple aggregative index number method and chain index number, so called a mixed method. A study case is given on the calculation of Microsoft's stock price index for 10 years i.e. 2007-2017 period.

The result of stock price index calculation by using simple aggregative method and stock price index for Microsoft stock then modeled with using binary logistic regression. Furthermore the two binary logistic regression models generated were compared. The comparison results showed that the binary logistic regression model of the mixed method is better than the simple aggregative method.

2. Calculation of the mixed index number

Generally, there are several methods to calculate the index number in the economic field. These methods can be categorized as chart following:



However, there also added a chain index method into chart as a own method between simple and weighted method.

This section also explains the method of calculating for simple aggregative and chain price index number, furthermore it results calculations of the mixed price index number. Calculation of simple aggregative price index number is done through the following formula:

$$SIN = \frac{\sum P_n}{\sum P_0} \, x \, 100 \qquad (1)$$

SIN : simple agregative price index number
$\sum P_n$: price on year for n
$\sum P_0$: price on the base year

Table 2.1 presents the stock prices and index numbers of Microsoft stock calculated by using the simple aggregative method for 10 years on 2007-2017 period.

Table 2.1 Stock price and index number of Microsoft stock in 2007-2017 period

Year	Price	Simple Aggregative Index Number
2007	2375018	100
2008	1463792	61,63288
2009	29458,46	1,240347
2010	4148422	174,6691
2011	2216547	93,32758
2012	3253521	136,9893

2013	1257098	52,93002
2014	946153,5	39,83773
2015	1084778	45,6745
2016	65928,29	2,775907
2017	9371,26	0,394576333

While the calculation of chain index numbers done with the formula:

$$CIN = \frac{P_i}{P_{i-1}} \, x \, 100 \quad (2)$$

CIN : chain index number
P_i : price on the year for i
P_{i-1}: price on the year for (i-1)

The stock price index can not be determined by using this chain method because the stock price in a given year is obtained from the total daily stock price for one year. However, it can be done by using a mixed method of simple aggregative and chain number index method by formula:

$$MIN = \frac{\sum P_i}{\sum P_{i-1}} \, x \, 100 \quad (3)$$

MIN : mixed index number
$\sum P_i$: price on the year for i
$\sum P_{i-1}$: price on the year for (i-1)

Table 2.2 presents the stock prices and index numbers calculated by using the mixed method for 10 years in 2007-2017 period.

Table 2.2 Stock price and index number of Microsoft stock in 2007-2017 period

Year	Price	Mixed Index Number
2007	2375018	100
2008	1463792	61,63288
2009	29458,46	2,012476
2010	4148422	14082,28
2011	2216547	53,43109
2012	3253521	146,7833
2013	1257098	38,63807
2014	946153,5	75,26491
2015	1084778	114,6514
2016	65928,29	6,077585
2017	9371,26	14,21432287

3. Logistic regression model and case study

Regression is a statistical model used to describe the relationship between a dependent variable and one or more independent variables. While logistic regression, independent variable is a discrete variable that has two or more possible values, if then the response variable is a discrete variable that has two values of probability so called binary logistic regression. Difference from logistic regression and linear regression is dependent variable in linear regression is a continuous variable, whereas in logistic regression is binary or dichotomous.

It is also explained that logistic regression is an approach to make predictive models as well as linear regression commonly or called Ordinary Least Square or OLS regression. It's difference, logistic regression predicts the dependent variable that has a dichotomy scale. The dichotomy is the nominal data scale with two categories,examples yes and no, good and bad, or high and low. It also

describe an approach to classification where some or all variables are qualitative variables, this approach is called logistic regression. OLS regression requires the requirement or assumption that error of variance or residual is normally distributed. Otherwise, logistic regression does not require the certain assumptions such as:

1. Logistic regression do not require a linear relationship between independent and dependent variables.
2. The independent variable does not require the multivariate normality assumptions.
3. The assumption of homocedasticity is not necessary.
4. The independent variable does not need to be in the form of interval or ratio scale.
5. Dependent variable must be dichotomy that is two categories.
6. The independent variable should not have the same diversity among the variable groups.
7. Categories in independent variables must be separate from each other or exclusive.
8. Samples required in relatively large quantities, minimum required up to 50 data samples for a independent variable.
9. Can select relationships because it uses a non-linear log transformation approach to predict the odds ratio. Odd in logistic regression is often expressed as a probability.

Logistic regression models are generally written as follows:

$$\pi(x) = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}} \quad (4)$$

with $\pi(x) = E(Y|x)$, i.e Y's expectation of x.

A transformation of $\pi(x)$ is a logit transformation defined in the form $\pi(x)$ is

$$g(x) = \ln\left(\frac{\pi(x)}{1-\pi(x)}\right) = \beta_0 + \beta_1 x \quad (5)$$

The importance of this transformation is that g (x) has properties of a linear regression.

Case study of logistic regression model can be applied to model the stock price index. Specifically, for modeling stock price index can use binary logistic regression by two categories i.e up and down. In this matter up is given by coding 1 and down is 0. Case study for Microsoft's stock price index numbers are as follows in table 2.3.

Table 2.3 presents price index numbers and categories to model Microsoft's stock price index numbers calculated by using simple aggregative method for 10 years in 2007-2017 period.

Table 2.3 Index numbers and categories of Microsoft stock in 2007-2017 period

Year	Price Index Number	Category
2007	100	base year
2008	61,63288	0
2009	1,240347	0
2010	174,6691	1
2011	93,32758	0
2012	136,9893	1
2013	52,93002	0
2014	39,83773	0
2015	45,6745	1
2016	2,775907	0
2017	0,394576333	0

Output on SPSS can be used to see the fitting model of Microsoft's stock price index number which calculation by using simple aggregative index number method on binary logistic regression model

from data in table 2.3. The results show that binary logistic regression model is not suitable to model simple aggregative index number of Microsoft stock price for 10 years in 2007-2017 period.

Variables in the Equation

		B	S.E.	Wald	Df	Sig.	Exp(B)
Step 1[a]	IHSATT	.813	232.854	.000	1	.997	2.254
	Constant	-93.375	26268.288	.000	1	.997	.000

a. Variable(s) entered on step 1: IHSATT.

Table 2.4 presents price index number to model stock price index numbers calculated by using the mixed method for 10 years in 2007-2017 period.

Table 2.4 Index numbers and categories of Microsoft stock in 2007-2017 period

Year	Price Index Number	Category
2007	100	base year
2008	61,63288	0
2009	2,012476	0
2010	14082,28	1
2011	53,43109	0
2012	146,7833	1
2013	38,63807	0
2014	75,26491	1
2015	114,6514	1
2016	6,077585	0
2017	14,21432287	1

Output on SPSS can be used to see the fitting model of Microsoft's stock price index number which calculations use the mixed method in binary logistic regression model of data in table 2.4. The results show that binary logistic regression model is not suitable to model index number of Microsoft stock price for 10 years in 2007-2017 period.

Variables in the Equation

		B	S.E.	Wald	Df	Sig.	Exp(B)
Step 1[a]	IHATTR	.035	.024	2.129	1	.145	1.036
	Constant	-2.213	1.549	2.040	1	.153	.109

a. Variable(s) entered on step 1: IHATTR.

4. Conclusion

Based on description above, it can be concluded that binary logistic regression model for Microsoft stock price index for 10 years in 2007-2017 period calculated by using simple aggregative index number method does not produce suitable model. However, if the two models are compared then the binary logistic regression model for the Microsoft stock price index for 10 years in 2007-2017 period calculated by using the mixed method is better than using a simple aggregative method.

ISBN: 978-979-3812-53-3

References

[1] Bishwapati, et al 2014 *Index Numbers* (India : The Istitute of Chartered Accountant of India (ICAI)) chapter 16

[2] Hosmer, David W, and Lemeshow S 2000 *Applied Logistic Regression Second Edition* (New York : John Wiley & Sons. Inc)

[3] Johnson, Richard A, and Wichern D W 2007 *Applied Multivariate Statistical Analysis Sixth Edition* (United Stated of America : Pearson Prentice Hall)

[4] Hidayat A 2015 *Interpretasi Regresi Logistik Dengan SPSS*. Retreived from https://www.Statistikian.com/2015/02