Predicting Stock Markets Using Binary Logistic Regression Based on Bry-Boschan Algorithm

Renanta Dzakiya Nafalana¹, Mujiati Dwi Kartikasari²
¹,²Universitas Islam Indonesia, Indonesia

ABSTRACT

In the stock market, there are bullish and bearish terms that are reflected in the movement of the stock price index. One of the stock price indexes listed on the Indonesia Stock Exchange (IDX) is the IDX Composite. Stock market conditions fluctuate along with changes in stock prices that move randomly, while investors expect market conditions to be active (bullish market). Several factors influence the movement of the IDX Composite, one of which is macroeconomic factors. The purpose of this research is to find out the condition of stock market as well as predict its condition using macroeconomic indicators. The method used to determine stock market conditions (bullish or bearish) is the Bry-Boschan algorithm, while the method used to predict the stock market using macroeconomic indicators is the binary logistic regression method. The Bry-Boschan algorithm is widely used to detect peaks and troughs in business cycle analysis. Binary logistic regression is used to model data with responses that have two categories or are in the form of binary numbers. Results show that the IDX Composite experienced 42 times (month) bearish periods and 191 times (month) experienced bullish periods. The obtained model has an accuracy value of 81.55%.

A. INTRODUCTION

In the stock market, there are terms bullish and bearish. Active market conditions (bullish market) occur due to price increases, which impact transaction volume. Passive market conditions (bearish market) occur due to price decreases followed by a decrease in transaction volume (Husnan, 2018). The ups and downs in stock prices, as indicated by the movement of the stock price index, are indicative of bullish and bearish market conditions (Novianto, 2011; Safitri, 2021). The IDX Composite, often known as the Indonesia Composite Index, is one of the stock price indices listed on the Indonesia Stock Exchange (IDX) (Sampurna et al., 2016). All equities listed on the Indonesia Stock Exchange are included in the IDX Composite index (Syarina, 2020).

Investors must have access to pertinent information on the state of the capital markets in order to make wise investment choices. Identification of stock market conditions is carried out to determine whether the stock market is active (bullish market) or passive (bearish market). The stock market conditions (bearish or bullish) can be identified by the Bry-Boschan algorithm. The Bry-Boschan algorithm is an approach widely employed in identifying the phases of business and financial cycles (Wu and Lee, 2015). The first study from Tüzen et al. (2022), with the aim of examining the fundamental characteristics of the cyclical fluctuations in the Turkish economy and to determine the business cycle (contraction and expansion) using Bry-Boschan (BB) algorithm. The research of Kaur (2020) with the aim of identifying the monetary sector’s leading indicators for the Indian economy. The research of Luvsannyaam et al. (2019), with the aim of comparing the Mongolian business cycle using graphical and parametric methodologies.

Stock market conditions fluctuate along with changes in stock prices that move randomly, while investors expect market condi-
tions to be an active or bullish market (Barus and Wijaya, 2021; Sasono, 2022; Triani, 2013). Prediction is used to find out how the condition of the stock market. As a result, investors may decide whether to increase or decrease their portfolio allocation in the stock market (Widoatmodjo, 2012).

Several factors influence the movement of the IDX Composite, one of which is macroeconomic factors (Blanchard, 2017). Macroeconomic factors that affect stock performance include international economic conditions, a country’s economic cycle, inflation rates, tax regulations, money supply, exchange rates, and interest rates for Bank Indonesia certificates (Krisna and Wirawati, 2013; Samsul, 2015). Several methods can be used to predict stock market conditions (bearish or bullish) for the IDX Composite, one of which is binary logistic regression. Binary logistic regression is a form of regression used to model the relationship between response variables and predictor variables, where a response variable is a binary number (Dewi and Pratiwi, 2021). As the research of Ali et al. (2018), with the aim of forecasting stock performance using a logistic regression model. The findings demonstrated that stock performance is highly predicted by financial and accounting parameters. For predicting positive or negative stock performance, the prediction accuracy is 89.77%.

Based on this, this research aims not only to identify stock market conditions (bearish or bullish) with the Bry-Boschan algorithm but also to predict them using macroeconomic indicators, namely: inflation, BI interest rates, the rupiah exchange rate against foreign currencies, especially the US dollar, and the amount of circulated money using the binary logistic regression method.

B. RESEARCH METHOD

The data used in this research is secondary data which contains monthly data from January 2003 to May 2022. The data is sourced from yahoo finance, the Central Bureau of Statistics (BPS) website, and the Bank Indonesia website.

This research uses a variable consisting of one dependent variable/response, the IDX Composite, and four independent variables/predictors, namely inflation, BI rates, the US dollar exchange rate, and the amount of circulated money. The method used in this research is the Bry-Boschan algorithm and binary logistic regression. Flowcharts of the research carried out are presented in Figure 1.

![Flowchart of the Research](image)

**Figure 1.** Flowchart of the Research

Explanation of the flowchart is as below.

1. Collect secondary data that is IDX Composite and macroeconomic indicators such as inflation, BI rates, US dollar exchange rates, and the amount of circulated money (M2).
3. Conduct a descriptive analysis of the IDX Composite variable to see a general description of the data.
4. Identify stock market conditions (bearish or bullish) IDX Composite uses the Bry-Boschan algorithm to find out when the IDX Composite is in a bearish period and when the IDX Composite is in a bullish period. The procedure for the Bry-Boschan algorithm is as follows:
   (a) The period with a higher or lower value than other values in the range of 5 months before and after identification is defined...
as a potential turning point or can be written as in Equation 1 and Equation 2. Its called a peak at time t if:

$$y_{t-k} < y_t and y_t > y_{t+k}$$  

(1)

Its called a trough if:

$$y_{t-k} < y_t and y_t < y_{t+k}$$  

(2)

with $k = 5$. $k$ is the minimum duration of an up or downtrend (Tüzenc et al., 2022).

(b) The turning point is called to be bearish or bullish if:

<table>
<thead>
<tr>
<th>Bearish</th>
<th>Bullish</th>
</tr>
</thead>
<tbody>
<tr>
<td>when the turning point was originally a peak, it turned into a trough.</td>
<td>when the turning point was originally a trough, it turned into a peak.</td>
</tr>
<tr>
<td>A peak to trough phase has a minimum duration of 5 months.</td>
<td>A trough to peak phase has a minimum duration of 5 months.</td>
</tr>
<tr>
<td>A peak to trough cycle has a minimum duration of 15 months.</td>
<td>A trough to peak cycle has a minimum duration of 15 months.</td>
</tr>
<tr>
<td>If there are two or more similar turning points (peak to peak) sequentially, the highest peak is selected.</td>
<td>If there are two or more similar turning points (trough to trough) sequentially, the lowest trough is selected.</td>
</tr>
<tr>
<td>If there are two or more turning points with the same value, then the last point is designated as a potential turning point</td>
<td>Turn points that occur within 6 months or less of the beginning and end of a data series period are not considered potential turning points.</td>
</tr>
</tbody>
</table>

5. Add macroeconomic variables, namely inflation, BI interest rates, US dollar exchange rate, and the amount of money in circulation (M2).

6. Conduct descriptive analysis of macroeconomic variables (inflation, BI interest rate, US dollar exchange rate, money supply) to see a general picture of the data.

7. Conduct data standardization

Data standardization formula:

$$Z = \frac{x - \bar{x}}{sd}$$  

(3)

with $Z$: Z-score, $x$: value from data, $\bar{x}$: mean data, $sd$: standard deviation

8. Conduct modeling with binary logistic regression analysis.

The logistic regression function is

$$\pi(x) = \frac{e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_j x_j}}{1 + e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_j x_j}}$$  

(4)

with $j$ is the number of independent or predictor variables in the model. $\pi(x)$ is the probability of success with a probability value $0 \leq \pi(x) \leq 1$ and $\beta$ is a parameter value. The logit form of this probability is:

$$g(x) = \ln \left( \frac{\pi(x)}{1 - \pi(x)} \right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_j x_j$$  

(5)

9. Conduct diagnostic tests.

(a) Overall test

The test statistic used is the G test statistic or Likelihood Ratio test as follows: (Hosmer Jr et al., 2013)

$$G = -2 \ln \left( \frac{n_0}{n} \frac{n_0}{n} \frac{n_1}{n} \frac{n_1}{n} \prod_{i=1}^{n} \frac{\hat{\pi}_y}{1 - \hat{\pi}_y} \right)$$  

(6)

with hypotheses $H_0 : \beta_1 = \beta_2 = \ldots = \beta_p = 0$, which means the predictor variable has no significant effect on the response variable. Reject $H_0$ if the $G$ test statistic value $> \chi^2_{(a,v)}$ or if the p-value $< \alpha$.

(b) Partial test

In binary logistics regression, the parameter significance test in the model can be carried out with the likelihood ratio test and the Wald test.

$$W^2 = -2 \frac{\hat{\beta}_j^2}{SE (\hat{\beta}_j)^2}$$  

(7)

with hypotheses $H_0 : \beta_j = 0, j = 1, 2, \ldots, p$ (The $j$-th predictor variable has no significant effect on the response variable). Wald test statistic follows the $\chi^2$ (Chi-Square) distribution. Reject $H_0$ if $W^2 > \chi^2_{(a,v)}$ or if p-value $< \alpha$. 

Vol. 6, No. 2, April 2023, Hal. 127–136
DOI: https://doi.org/10.30812/varian.v6i2.2385
Odds ratio is one of the measures used to interpret the coefficients of predictor variables. To determine the odds ratio can be seen in Equation 8,

\[ \text{OR} = \frac{\pi(1)/[1-\pi(1)]}{\pi(0)/[1-\pi(0)]} \]

it is known that \( \pi(1) = \frac{e^{(\beta_0 + \beta_j)}}{1 + e^{(\beta_0 + \beta_j)}} \) and \( \pi(0) = \frac{e^{(\beta_0)}}{1 + e^{(\beta_0)}} \) with \( j = 1, 2, \ldots, p \). The OR value is obtained in Equation 9

\[ \text{OR} = e^{\beta_j} \]

(d) Goodness of Fit Test

The goodness of fit test is done with the Hosmer and Lemeshow test which is carried out based on grouping the estimated probability values observed in each independent variable. The statistics of the Hosmer and Lemeshow test are stated in Equation 10

\[ \hat{C} = \sum_{k=1}^{g} \left( \frac{(O_k - n_k \tilde{\pi}_k)^2}{(n_k \tilde{\pi}_k (1 - \tilde{\pi}_k))^2} \right) \]

with \( g \) : number of groups, \( O_k \) : sum of the response variable values in the \( g \) group, \( \tilde{\pi}_k \) : mean estimated probability, \( n_k \) : number of observations in the \( g \) group. The hypotheses used are \( H_0 : \pi_k = \pi_0, k = 1, 2, \ldots, g \) (The model is suitable or there is no difference between the observation results and the predicted results). Reject \( H_0 \) if the test statistic \( \hat{C} value > \chi^2_{(\alpha, g-2)} \) or if the \( p-value < \alpha \).

(c) Confusion Matrix

A confusion Matrix is a measuring tool in the form of a matrix that is used to obtain the amount of classification accuracy of the class with the classification algorithm used (Witten et al., 2017). The response variable which has two classes has four possible results of different classification predictions, namely true positive (TP), true negative (TN), false positive (FP), and false negative (FN).

<table>
<thead>
<tr>
<th>Table 2. Confusion Matrix</th>
</tr>
</thead>
<tbody>
<tr>
<td>Predicted Positive</td>
</tr>
<tr>
<td>TP</td>
</tr>
<tr>
<td>FN</td>
</tr>
</tbody>
</table>

Prediction accuracy can be known by calculating the accuracy, precision, and specificity results. The confusion matrix table also shows the error results from the classification algorithm used. The following formula may be used to determine accuracy, precision, specificity, and error rate values

\[ \text{Accuracy} = \frac{TN + TP}{TP + FP + FN + TN} \]

\[ \text{Precision} = \frac{TP}{TP + FP} \]

\[ \text{Specificity} = \frac{TN}{TN + FP} \]

\[ \text{ErrorRate} = \frac{FP + FN}{TP + FP + FN + TN} \]

10. Conduct interpretation results.

C. RESULTS AND DISCUSSION

1. Identify Stock Market Conditions (Bearish or Bullish) IDX Composite with Bry-Boschan Algorithm

Before identifying the stock market conditions of the IDX Composite, a descriptive analysis was performed to find out the general picture of the data. As Figure 2, IDX Composite data from January 2003 to May 2022 fluctuated.
The condition of the IDX Composite for the period January 2003 to May 2022 was identified using the Bry-Boschan algorithm. A plot of the IDX Composite condition is obtained as shown in Figure 3.

Based on Figure 3, bearish periods are marked in dark or grey colours, while bullish are marked in light or white colours. In the period from January 2003 to May 2022, there were 5 bearish and 6 bullish. From the 1st year (2003) to the 5th year (2007) and in the 6th year (2008) to the 11th year (2013), the IDX composite experienced bullish. For details, see Table 4.

<table>
<thead>
<tr>
<th>Phase</th>
<th>Start</th>
<th>End</th>
<th>Duration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bullish</td>
<td>-</td>
<td>2007M12</td>
<td>-</td>
</tr>
<tr>
<td>Bearish</td>
<td>2007M12</td>
<td>2008M11</td>
<td>11</td>
</tr>
<tr>
<td>Bullish</td>
<td>2008M11</td>
<td>2013M5</td>
<td>54</td>
</tr>
<tr>
<td>Bearish</td>
<td>2013M5</td>
<td>2013M12</td>
<td>7</td>
</tr>
<tr>
<td>Bullish</td>
<td>2013M12</td>
<td>2015M3</td>
<td>15</td>
</tr>
<tr>
<td>Bearish</td>
<td>2015M3</td>
<td>2015M9</td>
<td>6</td>
</tr>
<tr>
<td>Bullish</td>
<td>2015M9</td>
<td>2018M10</td>
<td>29</td>
</tr>
<tr>
<td>Bearish</td>
<td>2018M2</td>
<td>2018M10</td>
<td>8</td>
</tr>
<tr>
<td>Bullish</td>
<td>2018M10</td>
<td>2019M7</td>
<td>9</td>
</tr>
<tr>
<td>Bearish</td>
<td>2019M7</td>
<td>2020M5</td>
<td>10</td>
</tr>
<tr>
<td>Bullish</td>
<td>2020M5</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
Based on Table 4, it is known that the bullish started from January 2003 to December 2007. The bearish started from December 2007 to November 2008 with a period duration of 11 months. The average duration of a bearish is 8.4 months and the average duration of a bullish is 26.8 months.

2. Prediction of Stock Market Conditions (Bearish or Bullish) IDX Composite Using Macroeconomic Indicators with the Binary Logistics Regression Method

After knowing the stock market conditions (bearish or bullish) on the IDX Composite, the next step is to predict the stock market conditions (bearish or bullish) on the IDX Composite using macroeconomic indicators. In this research, 4 macroeconomic indicators were used, namely inflation, BI rates, the US dollar exchange rate, and the amount of circulated money (M2).

Before starting the analysis using the binary logistic regression method, a descriptive analysis was carried out for each macroeconomic indicator. The results of the descriptive analysis on each macroeconomic indicator (inflation, BI interest rate, US dollar exchange rate, and the amount of circulated money) are presented in Figure 4.

Figure 4 shows a plot or graph of inflation and BI rates from 2003 to 2022. The movement of inflation tends to be more volatile compared to the BI rate. The peak or highest value of inflation and BI rates occurred in the same period, namely around the end of 2005 to early 2006. The average inflation from 2003 to 2022 was 5.72%, while the average BI Rate was 6.88%. Inflation has a maximum value of 18.38% and the lowest value of 1.32%. Meanwhile, the BI rate has a maximum value of 12.75% and a minimum value of 3.50%.

Based on the graph of the average annual dollar exchange rate in Figure 4, the price of the US dollar tends to increase from year to year. The dollar exchange rate experienced a significant increase from 2011-2015. The amount of circulated money (M2) from 2003 to 2022 reached 805.22 billion Rupiah. The amount of circulated money (M2) also increased from year to year.

1. Modeling with Binary Logistics Regression

Before modeling with binary logistic regression, standardization data is performed for macroeconomic variables (inflation, BI rates, dollar exchange rate, and the amount of circulated money) because these variables have different units.

Furthermore, binary logistic regression modeling was carried out and the model was obtained as follows: Probability model

\[ \pi(x) = \frac{e^{1.8445-1.7365x_1+1.6682x_2-2.0845x_3+1.4193x_4}}{1+e^{1.8445-1.7365x_1+1.6682x_2-2.0845x_3+1.4193x_4}} \]
Logit Model

\[ g(x) = \ln \left( \frac{\pi(x)}{1 - \pi(x)} \right) \]
\[ = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 \]
\[ = 1.8445 - 1.7365x_1 + 1.6682x_2 - 2.0845x_3 + 1.4193x_4 \]

2. Parameter Significance Test (Overall Test)

The overall test is used to determine whether the independent variables simultaneously or jointly affect the dependent variable so that the model obtained is feasible to use or not.

<table>
<thead>
<tr>
<th>Table 4. Overall Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>G</td>
</tr>
<tr>
<td>-----------</td>
</tr>
<tr>
<td>41.8757923</td>
</tr>
</tbody>
</table>

Based on Table 4, the \( G \) value is \( 41.8757923 > \chi^2_{(0.05;4)} \) value is 9.487729. Thus, with a 95% confidence level, the available data rejects \( H_0 \). It can be concluded that at least there is an independent or predictor variable that has a significant effect on the response variable.

3. Partial Test

In order to ascertain whether or not the model developed is practicable for application, the overall test is performed to assess if the independent variables simultaneously or jointly affect the dependent variable.

<table>
<thead>
<tr>
<th>Table 5. Partial Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wald</td>
</tr>
<tr>
<td>-----------</td>
</tr>
<tr>
<td>(Intercept)</td>
</tr>
<tr>
<td>Inflation</td>
</tr>
<tr>
<td>BI Rate</td>
</tr>
<tr>
<td>Dollar Exchange Rate</td>
</tr>
<tr>
<td>The amount of circulated money</td>
</tr>
</tbody>
</table>

Based on Table 5, Wald test statistics \( > \chi^2 \) or \( W^2 > \chi^2_{(v)} \). Thus, with a 95% confidence level, the available data rejects \( H_0 \). This means that predictor variables (inflation, BI rates, dollar exchange rate, and the amount of circulated money) have a significant effect on the response variable (IDX Composite).

4. Odds Ratio

Odds ratio is used to simplify the interpretation of binary logistic regression models. From the parameter significance test, both overall and partially, it’s known that the predictor variables that significantly affect the response variable are inflation, BI rates, dollar exchange rate, and the amount of circulated money.

<table>
<thead>
<tr>
<th>Table 6. Odds Ratio Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variables</td>
</tr>
<tr>
<td>--------------------------</td>
</tr>
<tr>
<td>Inflation</td>
</tr>
<tr>
<td>BI Rate</td>
</tr>
<tr>
<td>Dollar Exchange Rate</td>
</tr>
<tr>
<td>The amount of circulated money</td>
</tr>
</tbody>
</table>

Based on the odds ratio in Table 6, the binary logistic regression model can be interpreted as follows:

- Each increase in one unit of inflation will have an effect of 0.1761364 times greater for the occurrence of the bullish condition of IDX composite.
- Each increase of one unit of BI interest rate will have an effect of 5.3024533 times greater for the occurrence of the bullish condition of IDX composite.
- Each increase of one unit of the dollar exchange rate will have an effect of 0.1243685 times greater for the occurrence of the bullish condition of IDX composite.
- Every one unit increase in the money supply will have an effect of 4.1342299 times greater for the occurrence of the bullish condition of IDX composite.
5. Test Goodness of Fit

Goodness of fit test is carried out to evaluate the binary logistic regression model that has been formed using the Hosmer and Lemeshow test.

<table>
<thead>
<tr>
<th>Table 7. Odds Ratio Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hosmer and Lemeshow Test</td>
</tr>
<tr>
<td>--------------------------</td>
</tr>
<tr>
<td>13.682</td>
</tr>
</tbody>
</table>

Based on Table 7, the \( \hat{C} \) value is 13,682 < \( \chi^2_{(0.05;8)} \) value is 15.50731. Thus, with a 95% confidence level, the available data failed to reject \( H_0 \). This means that the model is appropriate (there is no difference between the observed results and the predicted results).

6. Confusion Matrix

The results of the accurate prediction of stock market conditions (bearish or bullish) IDX Composite with binary logistic regression are presented in Table 8.

<table>
<thead>
<tr>
<th>Table 8. Confusion Matrix</th>
</tr>
</thead>
<tbody>
<tr>
<td>Predicted</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>IDX Composite Conditions</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Total</td>
</tr>
</tbody>
</table>

Table 8 shows that from 233 data for the period from January 2003 to May 2022, there were 47 times (months) the IDX Composite experienced a bearish and 186 times (months) the IDX Composite experienced a bullish. Furthermore, based on Table 8, the accuracy, precision, specificity, and error rate are calculated as follows

\[
\text{Accuracy} = \frac{TN + TP}{TP + FP + FN + TN} = \frac{183 + 7}{233} = 0.8155 = 81.55\%
\]

\[
\text{Precision} = \frac{TP}{TP + FP} = \frac{7}{7 + 3} = 0.7 = 70\%
\]

\[
\text{Specificity} = \frac{TN}{TN + FP} = \frac{183}{183 + 3} = 0.9839 = 98.39\%
\]

\[
\text{ErrorRate} = \frac{FP + FN}{TP + FP + FN + TN} = \frac{3 + 40}{233} = 0.1845 = 18.45\%
\]

The accuracy value is 81.55%, the precision value is 70%, and the specificity value is 98.39%. In addition, the error rate value is 18.45%. Based on the accuracy value, this research is not only good at applying the Bry-Boschan algorithm to identify stock market conditions (bearish or bullish), but also good at predicting stock market conditions using macroeconomic indicators.

D. CONCLUSION AND SUGGESTION

This research was conducted in two cases. The first is to identify stock market conditions (bullish or bearish). The second is predicting stock market conditions with macroeconomic indicators. Based on the results of the identification of stock market conditions on the IDX Composite using the Bry-Boschan algorithm, from January 2003 to May 2022, there were 5 bearish and 6 bullish. The bearish trend started from December 2007 to November 2008, with a duration of 11 months. The average duration of a bearish position is 8.4 months, and the average duration of a bullish position is 26.8 months.

Based on the prediction of the IDX Composite stock market conditions using macroeconomic indicators (inflation, BI rates, US dollar exchange rates, and the amount of circulated money), the results show that from 233 data points for the period from January 2003 to May 2022, there were 47 times (months) the IDX Composite experienced a bearish trend and 186 times (months) the IDX Composite experienced a bullish trend. Therefore, we obtain an accuracy value of 81.55%, a precision value of 70%, and a specificity value of 98.39%. In addition, the error rate is 18.45%.
REFERENCES


