

Application of the ARIMAX Model in Assessing the Impact of Global Stock Price Index on Forecasting Indonesia's IHSG

Marzuki Marzuki^{1,*}, Muhammad Riswanda¹, Nurhasanah¹, Asep Rusyana¹
¹ Department of Statistics, Universitas Syiah Kuala, Banda Aceh 23111, Indonesia
*Corresponding email: marzuki@usk.ac.id

ARTICLE INFO	ABSTRACT
Article History Received : 22-03-2024 Revised : 22-04-2024 Accepted : 29-04-2024 Keyword ARIMAX; IHSG; Stocks; Time series analysis; Volatility;	Time series analysis can be classified into two parts when viewed based on the analysis data variables, namely univariate and multivariate time series analysis. The ARIMAX model is the development of the ARIMA model. The ARIMAX model is a multivariate time series analysis method consisting of exogenous and endogenous variables. This study aims to forecast the Composite Stock Price Index (IHSG) using the ARIMAX model, by looking at the influence of global stock price indices, namely the American stock price index (DJIA), Japanese stock price index (N225), and Chinese stock price index (SSEC). The results of the study show that the model used to forecast the 2019 IHSG is the ARIMAX model (4,1,4). The results of the 2019 IHSG forecast produce fluctuating data. The highest IHSG share price of 6,958.419 occurred in November, while the lowest share price of 6.591.566 occurred in January. Forecast accuracy measured using RMSE and MAPE obtained results of 144.5387 and 2.5121 respectively.
DOI: ...	©2024 TJoMA. All rights reserved.

1. INTRODUCTION

A time series is a series of observations of an event, incident, symptom, or change that occurs from time to time. Time series data is data collected based on time units such as hours, days, weeks, months, years, semesters, and data observed over time (Andiyasa et al., 2014). Time series analysis can be classified into two parts if viewed based on the analysis data variables, namely univariate and multivariate time series analysis. Univariate time series analysis only observes or uses one time series data variable, while multivariate analysis uses more than one time series data variable (Bodie et al., 2006).

The ARIMAX model is a multivariate time series analysis method consisting of exogenous and endogenous variables. Exogenous variables are variables that are considered to come from outside the model being measured. This variable is assumed to be not influenced by other variables in the model. Exogenous variables can influence the time series being predicted, but are not predicted by the model itself. These variables are added to the model to increase predictive ability by including additional information that might influence the behavior of the time series. Endogenous variables are variables that are predicted or explained by the model itself. The addition of exogenous variables gives the model more information to predict or explain variations in endogenous variables. The ARIMAX is used to explain changes in data and the relationship between exogenous variables and endogenous variables in a model. Suryani et al. (2018) have used the ARIMAX model to predict rainfall data in Semarang with sea surface temperature (SST) El-Nino as the exogenous variable.

This research uses time series data from the Indonesian composite stock price index (hereafter IHSG) and global stock price indices, namely the American stock price index Dow Jones Industrial Average (DJIA), the Japanese stock price index Nikkei 225 (N225), and the Chinese stock price index Shanghai Stocks Exchange Composite (SSEC) which is considered to have a large relationship or connection to the world economy and Indonesia. The American

(DJIA), Japanese (N225), and Chinese (SSEC) global stock indexes have an influence on the Indonesian stock price index, in the form of positive and negative influences (Indonesian Stock Exchange, 2010). The United States experienced a financial crisis in 2008-2009 which also had an influence on the world financial crisis, including countries in Asia, one of which was Indonesia. The crisis caused the composite stock price index (IHSG) to decline due to the fall in the American stock price index, namely DJIA. This condition shows that a country's global economy is linked between one country and other countries. A strong capital market can influence a weak capital market (Cools et al., 2009).

Financial markets have become increasingly connected globally. Events and changes in the global stock market can have a significant impact on a country's stock market, including IHSG. The global and local stock markets are linked, making analysis of the influence of the global stock price index important to understand the potential impact on the Indonesian IHSG. Stock market volatility may increase due to global events such as financial crises, changes in global interest rates, or economic uncertainty. Therefore, adopting the ARIMAX model can help predict IHSG by including exogenous variables, such as the global stock price index, to increase the accuracy of predictions in the face of conditions of volatility and uncertainty.

The ARIMAX model allows researchers to incorporate external factors, such as global stock price indices, into the auto regressive integrated moving average (ARIMA) model. This step can improve the model's ability to predict the IHSG by taking into account external factors that can influence the Indonesian stock market. This research has high relevance for several parties in the Indonesian capital market. More informed investment decisions can be taken by investors, financial analysts and decision makers by understanding the influence of global stock price indices and being able to respond better to global market changes. This research can provide a new contribution to the literature by applying the ARIMAX model to the context of the Indonesian stock market and identifying the impact of the global stock price index on the IHSG.

2. METHODS

This research uses secondary data obtained from the site www.finance.yahoo.com. The data analyzed in this research is monthly time series data of closing stock price index values for 4 countries for 9 years, namely from January 2010 to December 2018. The research involved 4 variables with 108 observations (months). The following are details of the variables used in this research:

Table 1. Research variable

No.	Symbol	Variable	Country
1	Y	IHSG	Indonesia
2	X_1	DJIA	USA
3	X_2	N225	Japan
4	X_3	SSEC	China

The data above was processed by following the following ARIMAX model forecasting procedure:

- 1) Get summary statistics.
- 2) Form a time series data plot.
- 3) Testing stationarity of variance and mean.
- 4) Identify the model by plotting the Auto Correlation Function (ACF) and Partial Auto Correlation Function (PACF) correlograms to obtain a tentative model.
- 5) Modeling data using the ARIMAX model.

- 6) Choose the best model based on the Akaike's Information Criteria (AIC) value. The model is considered relatively better if the AIC value is smaller.
- 7) Diagnose the best model from step 6 with 3 tests, namely white noise testing, error normality and parameter significance.
- 8) Measuring forecasting accuracy based on MAPE and RMSE values
- 9) Forecast IHSG data for the next 12 months and draw conclusions.

3. RESULT AND DISCUSSION

3.1.Descriptive Analysis

Summary statistics for IHSG, DJIA, N225, and SSEC data for 9 years, namely from 2010 to 2018, can be seen in the following Table 2.

Table 2. Statistics summary

Variable	Min	Average	Max
IHSG	2,549.03	4,702.50	6,605.63
DJIA	9,774.02	16,676.04	26,458.31
N225	8,434.61	15,188.97	24,120.04
SSEC	1,979.21	2,769.05	4,611.74

IHSG has an average share price of 4,702.54 with the lowest share price being 2,549.03 in February 2010, and the highest share price being 6,605.63 in January 2018. The DJIA index has an average share price of 16,676, 04 with the lowest share price being 9,774.02 in June 2010, and the highest share price being 26,458.31 in September 2018. The N225 Index has an average share price of 15,188.97 with the lowest share price being 8,434.61 in November 2011, and the highest share price was 24,120.04 in September 2018. The SSEC Index had an average share price of 2,776.38 with the lowest price being 1,979.21 in June 2013, and the highest share price being 4,611.74 in May 2015.



Figure 1. Plot of IHSG, DJIA, N225, and SSEC time series data

Data fluctuations for each variable can be seen in Figure 1. The Indonesian Composite Stock Price Index (IHSG) data plot has the form of a positive trend data pattern, with the data continuing to increase from year to year. The Dow Jones Industrial Average (DJIA) index data plot also has a positive trend data pattern, with DJIA index share prices also continuing to increase from year to year. The Nikkei 225 (N225) index data plot has stable data fluctuations from the beginning of 2010 to the end of 2012. However, at the beginning of 2013 the N225 index shows a positive trend until 2018. The Shanghai Stocks Composite Exchanges (SSEC) index data plot shows a negative trend, namely a decline from 2010 to 2014. In mid-2015,

precisely in April and May, there was a very significant sudden increase in the SSEC index, until it reached its highest point in May 2015. The SSEC index experienced another significant decline in the same year in July 2015. The extreme movement of the Chinese stock index was caused by the turmoil in the Chinese stock market that year. At that time, China's economic growth was slowing or weakening, which was inversely proportional to the Chinese stock market index which was increasing rapidly.

3.2. ARIMAX Model

Stationarity testing in variance was carried out using Lambda BoxCox. Table 3 presents the lambda values. The lambda value can be estimated using a likelihood function.

Table 3. Lambda Box-Cox value

Variable	Lambda (λ) value
IHSG (Y)	0.315
DJIA (X_1)	-0.173
N225 (X_2)	-0.083
SSEC (X_3)	-0.447

The results show that all variables have a lambda value $\lambda \neq 1$ or not yet close to one. So it can be concluded that the data is not stationary regarding variance, so it is necessary to carry out transformations on the data. The transformation is carried out by powering the data as shown in Table 4. The power is obtained through trial and error until the lambda value $\lambda = 1$ or close to 1 is obtained. The type of transformation and the lambda value obtained are presented in Table 4.

Table 4. Box-Cox lambda values after transformation

Variable	transformation	Lambda (λ) value
IHSG (Y)	($y^{0.31}$)	1.027
DJIA (X_1)	($x_1^{-0.17}$)	1.007
N225 (X_2)	($x_2^{-0.11}$)	0.986
SSEC (X_3)	($x_3^{-0.11}$)	1.019

The lambda (λ) value of the four types of transformation is close to 1. All variables are stationary with respect to variance. Table 5 shows the results of the stationarity test on the mean. The four variables used in the research are not stationary in mean. This can be seen from the p-value $>$ alpha ($0.071 > 0.05$), ($0.323 > 0.05$), ($0.597 > 0.05$), ($0.577 > 0.05$). So the decision is that we cannot reject H_0 , meaning the data is not stationary in the mean. So differencing is needed in the data. The following are the results obtained after differencing the data.

Table 5. ADF test statistical value

Variable	Before differencing		After differencing	
	t-count	p-value	t-count	p-value
IHSG (Y)	-3.323	0.071	-5.046	0.01
DJIA (X_1)	-2.613	0.323	-5.425	0.01
N225 (X_2)	-1.952	0.597	-3.876	0.01
SSEC (X_3)	-2.000	0.577	-4.282	0.01

3.3. Model Identification

Model identification to determine the order of the ARIMA model is done by looking at the pattern of the ACF and PACF plots from stationary data. The data used to view the ACF and PACF plots uses IHSG data which is an endogenous variable. The plot of the ACF and PACF plots are as follows:

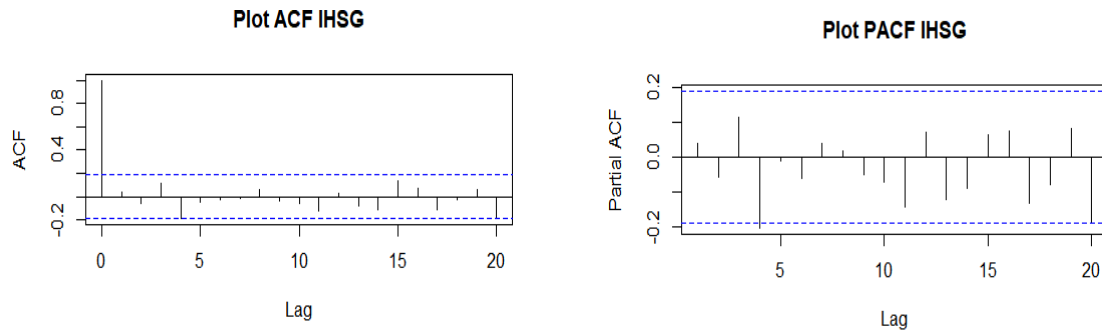


Figure 2. Plot ACF and PACF of IHSG data

Figure 2 shows that the ACF plot pattern of the IHSG data decreases rapidly after lag 0, and cuts off at the second lag. And the pattern of the PACF plot is truncated after the first lag. Based on the ACF plot, it can be seen that there is only one lag that crosses the line, namely the fourth lag. And in the PACF plot there is also one lag that crosses the line, namely also the fourth lag. Based on the results of the ACF and PACF plots, it is suspected that the appropriate ARIMA model for this data is the ARIMA model (4,1,4) with one differencing of the data. Because there is only one presumptive model, the ARIMAX model modeling starts with the tentative ARIMAX model (1,1,1), which is continued by selecting the best model from the 10 models that have been formed.

Table 6. ARIMAX model and AIC value

Model	Order	AIC
ARIMAX 1	(1,1,1)	1401.26
ARIMAX 2	(2,1,1)	1401.59
ARIMAX 3	(1,1,2)	1402.89
ARIMAX 4	(2,1,2)	1401.89
ARIMAX 5	(3,1,2)	1403.84
ARIMAX 6	(2,1,3)	1401.81
ARIMAX 7	(3,1,3)	1405.98
ARIMAX 8	(4,1,3)	1405.63
ARIMAX 9	(3,1,4)	1405.77
ARIMAX 10	(4,1,4)	1401.19

Table 6 is the AIC value of the 10 models that have been tried. Of the 10 models, there are 5 models that have the smallest AIC values, with the differences in AIC values not being much different or significant from each other, namely the ARIMAX 1, 2, 4, 6 and 10 models.

3.4. Model Diagnosis

White noise testing was carried out using the Ljung-Box test. Table 7 shows the results of the white noise test. These five models have fulfilled the white noise assumption. This order can be seen from the p-value produced by ARIMAX 1, 2, 4, 6 and 10 which is greater than $\alpha = 0.05$. Respectively, namely $(0.875 > 0.05)$, $(0.646 > 0.05)$, $(0.728 > 0.05)$, $(0.910 > 0.05)$,

(0.745 > 0.05). So the decision is that we cannot reject H_0 , meaning that the five models do not have autocorrelation between errors.

Table 7. Statistical results of the Ljung-Box test on the ARIMAX model (p,d,q)

Model	χ^2	p-value
ARIMAX 1 (1,1,1)	0.025	0.875
ARIMAX 2 (2,1,1)	0.212	0.646
ARIMAX 4 (2,1,2)	0.121	0.728
ARIMAX 6 (2,1,3)	0.013	0.910
ARIMAX 10 (4,1,4)	0.106	0.745

The error normality test was carried out using the Shapiro Wilk test. Table 8 is the test results.

Table 8. ARIMAX model error normality test statistics (p,d,q)

Model	W	p-value
ARIMAX 1 (1,1,1)	0.966	0.007
ARIMAX 2 (2,1,1)	0.970	0.016
ARIMAX 4 (2,1,2)	0.976	0.046
ARIMAX 6 (2,1,3)	0.985	0.136
ARIMAX 10 (4,1,4)	0.991	0.700

Table 8 shows that the ARIMAX 6 (2,1,3) and ARIMAX 10 (4,1,4) models are normally distributed models, this is because they have a p-value > $\alpha = 0.05$, namely (0.136 > 0.05), (0.700 > 0.05). So the decision is that we cannot reject the null hypothesis. And it can be concluded that the ARIMAX 6 and 10 models have normal distribution errors. And the other three models, namely ARIMAX models 1, 2, and 4, produce p-values < α , respectively (0.007 < 0.05), (0.016 < 0.05), and (0.046 < 0.05). So the decision is to reject the null hypothesis. And it can be concluded that the error does not spread normally.

Parameter significance testing is used to determine whether the influence of a parameter in the model is significant or not. A parameter is said to have a significant influence if the p-value < $\alpha = 0.05$. Based on the results of previous tests, there are two models that meet the white noise and error normality assumptions, namely the ARIMAX 6 (2,1,3) and ARIMAX 10 (4,1,4) models. So it can be concluded that the two models are suitable for use for forecasting. Test statistics for testing parameter significance using equation (2.14) obtained results as in Table 9.

Table 9. Estimation of ARIMAX model parameters (2,1,3)

Coefficient	Estimated Value	p-value	Significance
ϕ_1	0.664	0.000	Significant
ϕ_2	-0.895	0.000	Significant
θ_1	-0.630	0.000	Significant
θ_2	1.022	0.000	Significant
θ_3	-0.037	0.767	No
β_1	0.145	0.000	Significant
β_2	-0.025	0.342	No
β_3	0.095	0.293	No

Table 9 shows that there are three parameters that are not significant in the results of parameter significance testing in the ARIMAX model (2,1,3). This result can be seen from the p-value of these three parameters being greater than $\alpha = 0.05$. The equation for the ARIMAX model (2,1,3) can be written as follows:

$$(1-B)^1 (1+0,6637B-0,8950B^2) Y_t = (1-0,6299B+1,0219B^2+0,0371B^3) e_t + 0,1446X_{1,t} - 0,0252X_{2,t} + 0,0949X_{3,t} \quad (1)$$

The following are the results of testing the significance of the ARIMAX (4,1,4) model parameters:

Table 10. Parameter estimation of the ARIMAX (4,1,4) model

Coefficient	Estimated Value	p-value	Significance
ϕ_1	1.069	0.000	Significant
ϕ_2	-0.365	0.013	Significant
ϕ_3	0.965	0.000	Significant
ϕ_4	-0.819	0.017	Significant
θ_1	-1.207	0.021	Significant
θ_2	0.473	0.000	Significant
θ_3	-1.207	0.000	Significant
θ_4	1.000	0.000	Significant
β_1	0.212	0.000	Significant
β_2	-0.051	0.040	Significant
β_3	-0.044	0.548	No

Table 10 shows that almost all parameters have a significant effect in the model. Except for the parameters of the SSEC stock index, which is a Chinese state stock. The decision can be seen from the resulting p-value of $0.548 > \alpha = 0.05$, so it cannot reject the null hypothesis, and the parameter is not significant. The equation for the ARIMAX model (4,1,4) can be written as follows:

$$(1-B)^1 (1+1,0678B-0,3646B^2+0,9653B^3+0,8194B^4) Y_t = (1-1,2069B+0,4727B^2+1,2069B^3+1,0000B) e_t + 0,2116X_{1,t} - 0,0508X_{2,t} - 0,0435X_{3,t} \quad (2)$$

The results of parameter significance testing from the two models above, the ARIMAX model (4,1,4) has only one insignificant parameter, and in the ARIMAX model (2,1,3) there are three insignificant parameters. So it can be concluded that the ARIMAX model (4,1,4) is better used for forecasting. In addition, the forecasting error value generated from the ARIMAX model (4,1,4) is smaller than the error value from the ARIMAX model (2,1,3).

3.5. Forecasting Accuracy Measurement and Data Forecasting

The measuring instruments used to calculate forecasting accuracy are mean absolute percentage error (MAPE) and root of mean square error (RMSE). The smaller the MAPE and RMSE values, the better the forecast results. From the results of diagnostic tests, there are two models that meet the assumptions, namely ARIMAX (2, 1, 3), and ARIMAX (4, 1, 4). Based on the results of forecasting accuracy measurements obtained, forecasting using the ARIMAX model (4,1,4) has MAPE and RMSE values smaller than the ARIMAX model (2,1,3) as seen in Table 11.

Table 11. MAPE and RMSE values

Model	RMSE	MAPE
ARIMAX (2,1,3)	151,374	2,643
ARIMAX (4,1,4)	144,539	2,512

Table 11 shows that forecasting using the ARIMAX model (4,1,4) produces better forecasts than using the ARIMAX model (2,1,3). IHSG 2019 data forecasting uses the ARIMAX model (4,1,4) by looking at the influence of the global stock price index, the resulting forecast is as in Table 11. Based on the forecast results, the highest share price of 6,958,419 occurred in November, and the lowest share price of 6,591,566 occurred in January.

Table 12. IHSG data forecast for 2019

Mounth	Forecast Data	Actual Data
January	6,591.57	6,197.87
February	6,777.02	6,552.06
March	6,788.14	6,468.62
April	6,908.24	6,485.72
May	6,623.36	6,458.12
June	6,924.87	6,277.29
July	6,956.26	6,281.18
August	6,873.80	6,385.26
September	6,866.52	6,331.15
October	6,795.66	6,163.98
November	6,958.42	6,225.81
December	6,856.79	6,299.54

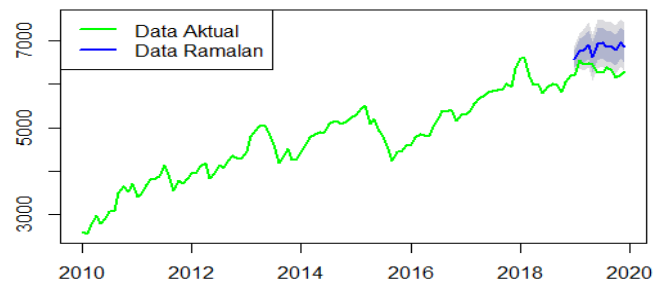


Figure 3. Plot data forecast results

Figure 3 shows IHSG forecast data for 2019. The results of the forecast can be seen on the blue line, IHSG experiences data fluctuations, which tends to show an upward trend in 2019.

4. CONCLUSION

The conclusion of this study is that the model used to forecast IHSG for 2019 is the ARIMAX model (4,1,4), with the following equation:

$$(1 - B)^1(1 + 1,068B - 0,365B^2 + 0,965B^3 + 0,819B^4) Y_t = (1 - 1,207B + 0,473B^2 + 1,207B^3 + 1,000B) e_t + 0,212X_{1,t} - 0,051X_{2,t} - 0,044X_{3,t}$$

The ARIMAX model (4,1,4) has better forecasting accuracy than other models, measured using MAPE and RMSE values of 2.512 and 144.539 respectively. The results of the IHSG data forecast for 2019 produce fluctuating data. The highest share price of 6,958,419 occurred in November, while the lowest share price of 6,591,566 occurred in January.

REFERENCES

Andiyasa, I. G. A., Purbawangsa, I. B. A., & Rahyuda, H. (2014). Pengaruh beberapa indeks saham dan indikator ekonomi global terhadap kondisi pasar modal Indonesia. *E-Jurnal Ekonomi dan Bisnis Universitas Udayana*, 44686.

- Bodie, Z., Kane, A., & Marcus, A.J. (2006). *Investments*, Edisi Keenam. Salemba Empat, Jakarta.
- Box, G. E., & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 26(2), 211-243.
- BEI. (2010). Buku panduan indeks harga saham bursa efek Indonesia. *Jakarta: Bursa Efek Indonesia*.
- Cools, M., Moons, E., & Wets, G. (2009). Investigating the variability in daily traffic counts through use of ARIMAX and SARIMAX models: assessing the effect of holidays on two site locations. *Transportation research record*, 2136(1), 57-66.
- Damodar, N. (2004). Basic econometrics: Student solutions manual for use with basic econometrics.-4th.
- Fakhruddin, H. M., & Darmadji, T. (2011). Pasar Modal di Indonesia. *Edisi Ketiga, Penerbit Salemba Empat*.
- Makridakis, S., Wheelwright, S. C., McGee, V. E., Andriyanto, U. S., & Basith, A. (1999). Metode dan Aplikasi Peramalan Jilid 1 Edisi Kedua. *Terjemahan oleh Ir. Hari Sumito. Jakarta: Bina Rupa Aksara*.
- Montgomery, D. C., Jennings, C. L., & Kulahci, M. (2015). *Introduction to time series analysis and forecasting*. John Wiley & Sons.
- Mulyana, D. (2004). *Metodologi penelitian kualitatif: paradigma baru ilmu komunikasi dan ilmu sosial lainnya*. PT. Remaja Rosdakarya.
- Nachrowi, D. N., & Usman, H. (2006). Ekonometrika: Pendekatan Populer dan Praktis Untuk Analisis Ekonomi dan Keuangan. *Jakarta: LPFE UI*.
- Oktarina, D. (2015). Pengaruh beberapa indeks saham global dan indikator makroekonomi terhadap pergerakan IHSG. *Journal of Business & Banking*, 5(2), 163-182.
- Sungkawa, I., & Megasari, R. T. (2011). Penerapan ukuran ketepatan nilai ramalan data deret waktu dalam seleksi model peramalan volume penjualan pt satriamandiri citramulia. *ComTech: Computer, Mathematics and Engineering Applications*, 2(2), 636-645.
- WEI, W. W. (2006). Univariate and Multivariate Methods. *TIME SERIES ANALYSIS*.
- Yaffee, R., & McGee, M. (2000). Introduction to Time Series Analysis and Forecasting: With Applications of SAS and SPSS Academic Press. *Inc, Orlando, FL, USA*.