Integra: Journal of Integrated Mathematics and Computer Science
Journal Homepage: https://integrajimcs.com/index.php/integra/index
integra
Journal of Integrated Mathematics and Computer Science

Research Paper

Application of State Space Representation on Vector Autoregressive (Var) Models for Forecasting

Oktaviana, S.[1], Widiarti[1*], Mustofa Usman[1], Edwin Russel[1], Jamal I. Daoud[2]

[1] *Faculty of Mathematics and Natural Science, Universitas Lampung, Bandar Lampung, 35141, Indonesia*
[2] *International Islamic University Malaysia, Selangor, 53100, Malaysia*
[*] **Corresponding author**: widiarti.1980@fmipa.unila.ac.id

Keywords
State Space, Vector Autoregressive, State Vector, Granger Causality Test, Forecasting

Abstract
Various analytical techniques are available for modeling multivariate time series data. One such approach is the State Space Model, which can be employed to model this type of data. In this study, the data to be analyzed are data on the Indonesia Rupiah (IDR) exchange rate (ExR) against the US Dollar (USD), oil and gas exports (OGE), money supply (MS) and non-oil and gas exports (non-OGE) from January 2008 to December 2019. The aim of this study is to identify the most suitable state space model for the given data. In this research, the state space method will be applied to multivariate time series data, with the state space represented in the Vector Autoregressive (VAR) model to explore the interrelationships among groups of observed variables. The VAR model is a statistical technique used to analyze the relationships between variables in the dataset, employing the Granger Causality Test. The state space model is utilized to model and forecast multiple interconnected time series, where the variables exhibit dynamic interactions and to examine additional unobserved variables in the time series data. Based on the analysis results and the minimum value of the Akaike Information Criterion (AIC), the optimal VAR model identified is the VAR (6) model. The results of forecasting values using the state space model show that the predicted values and the real values for the state space model are very closed to each other.

Received: 8 September 2025, Accepted: 11 March 2026
https://doi.org/10.26554/integrajimcs.20263141

1. INTRODUCTION

There are many approaches have been discussed in the last fifty years for analyzing data time series, either univariate or multivariate time series [1, 2, 3, 4, 5, 6, 7, 8]. The application of those methods for analysis data time series have been applied in many fields of science such as in economic, business, financial, biology, environments. One of the models that can be used to analysis multivariate time series data is state space model. This model was first introduced by Kalman [9] and further developed Kalman and Bucy [10], and it is widely used for modeling and forecasting multivariate time series with dynamic interactions. Considering the autocorrelations across the entire set of variables, the state space model can provide more accurate forecasts compared to methods that model each time series individually. In recent years, state-space representations and the corresponding Kalman recursions have significantly influenced time series analysis and various related fields. These techniques were initially developed for the control of linear systems (for accounts of this subject, see the books of Davis and Vinter [11] and Hannan and Deistler [12]). State space modeling provides a unified methodology for analyzing many problems in time series data [8].

The state space method was first developed in the field of control engineering, beginning with the pioneering paper by Kalman [9]. Initially, this approach was employed to precisely track the position and velocity of moving objects, such as ships, aircraft, missiles, and rockets. Around the seventies, scientists in other fields also recognized that this idea could also be applied to time series analysis in general [13, 14]. Since then the state space method has been applied in a variety of subjects, including business, biology, economics, finance, political science, environmental science, to medical science. The modern development of state space models began with the works of Kalman [9] and Kalman and Bucy [10], leading to a significant body of literature in engineering [15, 16]. Early statistical contributions included the Markovian representation proposed by Akaike [13, 14]. Hannan and Deistler [12] offered a comprehensive synthesis of the engineering and statistical time series research for stationary time series. In economics, Aoki and Havenner [17] explored multivariate state space models and proposed methodologies for both stationary and nonstationary data. For a detailed review of literature in this field, see Durbin and Koopman [8]. Over the past decade, state space modeling has become widely adopted in economics and finance. Key textbook treatments of state space models can be found in [6, 8, 18, 19, 20, 21, 22, 23, 24]. Gomez [24] discussed extensively about the state space model for multivariate time series data. At present, the state space method is used in statistics for analysis data time series. According to Wei [5], the state space model is a technique for modeling and simultaneously forecasting multiple interconnected time series data variables, where these variables exhibit dynamic interactions. The primary objective of the state space method is to derive meaningful results from a series of vectors based on the observed data.

The state space method is flexible and can be applied to both univariate and multivariate data. For extensive discussion of the application of state space in multivariate time series can be seen in Gomez [24]. While it is often used with single-variable data that does not require relationships between observed variables, this study will extend the application of the state space method to multivariate data. The state space will be represented using the Vector Autoregressive (VAR) model to examine the interrelationships among observed groups of variables. The VAR model is a statistical approach that can be utilized to explain the relationships between variables within the data.

Akaike [25, 26] introduced state vectors as vectors of canonical variates that relate the data to future observations in the development of state space models for time series. His approach can be seen as a generalization of a similar concept used for two vectors in factor models. This generalization is significant because, unlike static factor models, dynamic factors cannot be a mixture of past data and unobserved future realizations of the data vectors. Wei [5] explored the use of canonical correlation for fitting state space models. According to Durbin [27], state space models offer a powerful framework for practical time series analysis across various fields, including statistics, econometrics, and engineering.

The aim of this study is to identify the state space model that best fits the data multivariate time series data: the Indonesia Rupiah (IDR) exchange rate (ExR) against the US Dollar (USD), oil and gas exports (OGE), money supply (MS), and non-oil and gas exports (non-OGE). Then this model will be used to explore the behavior of the data such as the Granger Causality among data, the plot of the predicted values and the real data whether they are closed to each other or not, and finally the model will be used for forecasting.

2. METHODS

The data used in this study are data on the IDR exchange rate against the USD, oil and gas exports, money supply and non-oil and gas exports from January 2008 to December 2019. The data source was obtained from the Ministry of Trade and Bank Indonesia website.

The first step in analyzing time series data is to verify the stationarity assumption, which is crucial in time series analysis [2, 3, 28, 29, 30, 31]. In this study, the stationarity of the time series data is assessed both visually, by examining the data plot, and through the Augmented Dickey-Fuller (ADF) test. To test the stationarity of the time series data using the ADF test, the following model can be applied:

$$\Delta z_t = \mu + \beta t + \delta z_{t-1} + \sum_{i=1}^{m} \alpha_i \Delta z_{t-i} + e_t \tag{1}$$

The null and the alternative hypotheses are defined as follows:

$$H_0 : \delta = 0 \qquad \text{and} \qquad H_1 : \delta < 0$$

To test the null hypothesis, we use either the test-τ or the Dickey–Fuller test as follows:

$$\tau = \frac{\delta}{s_\delta} \tag{2}$$

The null hypothesis is rejected if the p-value $\leq \alpha$, for $\alpha = 0.05$ [28, 32, 33].

2.1 Model Vector Autoregressive (VAR)

As stated by Tsay [7], when analyzing multivariate time series data that involves multiple variables, one possible model to use is the Vector Autoregressive (VAR) model. The VAR(p) model can be expressed as follows:

$$\mathbf{Z}_t = \mathbf{\Phi}_0 + \sum_{i=1}^{p} \mathbf{\Phi}_i \mathbf{Z}_{t-i} + \mathbf{a}_t \tag{3}$$

where $\mathbf{Z}_t$ is an $n \times 1$ vector time series, $\mathbf{\Phi}_0$ is an $n \times 1$ vector of constants, $\mathbf{\Phi}_i$ is an $n \times n$ parameter matrix ($i > 0$, $\mathbf{\Phi}_p \neq 0$), and $\mathbf{a}_t$ is a shock vector with mean vector $\mathbf{0}$ and variance–covariance matrix Σ_a.

2.2 Granger Causality Test

The Granger Causality Test is one of the most widely used methods for examining causal relationships in multivariate time series data studies. According to [6, 34], this test helps to identify the short-term reciprocal relationships between the variables being studied. For example, to analyze the Granger causality between variables X and Y, the model for the Granger Causality Test is as follows:

$$x_t = c_1+\alpha_1 x_{t-1}+\alpha_2 x_{t-2}+\cdots+\alpha_p x_{t-p}+\beta_1 y_{t-1}+\beta_2 y_{t-2}+\cdots+\beta_p y_{t-p}+u_t \tag{4}$$

Under the assumptions of Ordinary Least Squares (OLS), the null hypothesis to be tested is stated as follows:

$$H_0 : \beta_1 = \beta_2 = \cdots = \beta_p = 0$$

(Y does not Granger-cause X)

against

$$H_1 : \text{at least one } \beta_p \neq 0$$

(Y Granger-causes X)

The statistical test is formulated as follows:

$$F = \frac{(RSS_0 - RSS_1)/p}{RSS_1/(T-2p-1)} \tag{5}$$

Reject the null hypothesis if F-Test $> F(\alpha, p, T-2p-1)$ or if p-value < 0.05 [?]. To calculate the residual sum of squares 1 (RSS_1), the shocks of model (4) are calculated as follows:

$$RSS_1 = \sum_{t=1}^{T} \hat{u}_t^2 \tag{6}$$

Under the null hypothesis, model (4) is represented as follows:

$$x_t = c_0 + \gamma_1 x_{t-1} + \gamma_2 x_{t-2} + \cdots + \gamma_p x_{t-p} + e_t \tag{7}$$

Next, to calculate the residual sum of squares 0 (RSS_0), the shocks of model (7) are calculated as follows:

$$RSS_0 = \sum_{t=1}^{T} \hat{e}_t^2 \tag{8}$$

2.3 Model State Space

The state space model is a method used to simultaneously model and predict multiple interconnected time series data, where the variables exhibit dynamic interactions [4]. The state space representation is explained through the state equation:

$$Y_{t+1} = AY_t + GX_{t+1} \tag{9}$$

And the output equation is:

$$Z_t = HY_t \tag{10}$$

where Y_t is a $k \times 1$ state vector, A is a $k \times k$ transition matrix, G is a $k \times n$ input matrix, X_t is an $n \times 1$ input vector, Z_t is an $m \times 1$ output vector, and H is an $m \times k$ observation (output) matrix [4].

2.4 Canonical Correlation Analysis

Canonical correlation analysis is a statistical method used to examine the relationship between a set of dependent variables and a set of independent variables. According to Wei [5], the state vector is uniquely defined through the analysis of canonical correlations between a series of current and past observations and a set of observations of present and future events. For a detailed discussion of canonical correlation, refer to Wei [5]. Canonical correlation analysis is conducted between data spaces as follows:

$$D_n = (Z'_n, Z'_{n-1}, \ldots, Z'_{n-p})' \tag{11}$$

and the predictor space:

$$F_n = (Z'_n, Z'_{n+1|n}, \ldots, Z'_{n+p|n})' \tag{12}$$

The order p for the VAR model is selected based on the optimal fit of the data to the VAR(p) model, which is determined by the smallest value of the Akaike Information Criterion (AIC). Canonical correlation analysis is applied to the Block Hankel matrix of the sample covariance:

$$\hat{\Gamma} = \begin{pmatrix} \hat{\Gamma}(0) & \hat{\Gamma}(1) & \cdots & \hat{\Gamma}(p) \\ \hat{\Gamma}(1) & \hat{\Gamma}(2) & \cdots & \hat{\Gamma}(p+1) \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\Gamma}(p) & \hat{\Gamma}(p+1) & \cdots & \hat{\Gamma}(2p) \end{pmatrix} \tag{13}$$

where $\hat{\Gamma}(j)$, $j = 0, 1, 2, \ldots, 2p$ is the sample covariance matrix defined as follows:

$$\hat{\Gamma}(s) = \frac{1}{n}\sum_{t=1}^{n-s}(Z_t - \bar{Z})(Z_{t+s} - \bar{Z})' \tag{14}$$

The components of the state vector are selected using the smallest canonical correlation of $\Gamma(j)$ [4].

2.5 Parameter Estimation

Based on the state space model

$$Y_{t+1} = AY_t + GX_{t+1}$$

where X_t – iid $N(0, \Sigma)$. The observation vector Z_t consists of m main components of Y_t so that $Z_t = HY_t$, where $H = [I_m \; 0]$ is a matrix of order $m \times k$.

As stated by Wei [5], once the state space model is selected, the next step is to estimate the parameters of the model using the maximum likelihood estimation method. Given the likelihood function:

$$L(Z_1, Z_2, \ldots, Z_n \mid A, G, \Sigma) = (2\pi)^{-\frac{n}{2}}\Sigma^{-\frac{n}{2}}e^{-\frac{1}{2}\,\text{trace}\, S(A,G)} \tag{15}$$

Table 1. Augmented Dickey Fuller Test for Data IDR Exchange Rate Against the USD (ExR), Oil and Gas Exports (OGE), Money Supply (MS) and Non-oil and Gas Exports (non-OGE)

Dickey-Fuller Unit Root Test					
Variable	Type	Rho	Pr<Rho	Tau	Pr<Tau
Kurs	Zero Mean	0.36	0.7682	1.02	0.9192
	Single Mean	-1.76	0.8053	-0.89	0.7896
	Trend	-8.44	0.5384	-2.05	0.5696
EM	Zero Mean	-1.19	0.4399	-0.98	0.2915
	Single Mean	-4.57	0.4726	-1.38	0.5888
	Trend	-8.18	0.5589	-2.03	0.5808
UB	Zero Mean	1.16	0.9333	4.03	0.9999
	Single Mean	0.45	0.9748	0.57	0.9884
	Trend	-20.13	0.0582	-3.23	0.0836
ENM	Zero Mean	0.19	0.7258	0.30	0.7714
	Single Mean	-9.56	0.1432	-2.32	0.1670
	Trend	-13.33	0.2343	-2.59	0.2842

Table 2. Augmented Dickey Fuller Test for Data IDR Exchange Rate Against the USD (ExR), Oil and Gas Exports (OGE), Money Supply (MS) and Non-oil and Gas Exports (non-OGE) After Differencing ($d = 1$)

Dickey-Fuller Unit Root Test					
Variable	Type	Rho	Pr<Rho	Tau	Pr<Tau
Kurs	Zero Mean	-171.74	0.0001	-9.20	<.0001
	Single Mean	-177.77	0.0001	-9.33	<.0001
	Trend	-177.85	0.0001	-9.29	<.0001
EM	Zero Mean	-182.19	0.0001	-9.55	<.0001
	Single Mean	-183.28	0.0001	-9.55	<.0001
	Trend	-183.84	0.0001	-9.52	<.0001
UB	Zero Mean	-191.93	0.0001	-9.63	<.0001
	Single Mean	-275.13	0.0001	-11.53	<.0001
	Trend	-280.91	0.0001	-11.62	<.0001
ENM	Zero Mean	-308.80	0.0001	-12.33	<.0001
	Single Mean	-311.30	0.0001	-12.33	<.0001
	Trend	-311.75	0.0001	-12.30	<.0001

So that the log likelihood function is:

$$\ln L(Z_1, Z_2, \ldots, Z_n \mid A, G, \Sigma) = \ln \left[(2\pi)^{-\frac{n}{2}} \Sigma^{-\frac{n}{2}} e^{-\frac{1}{2}\operatorname{trace} S(A,G)} \right]$$
$$\ln L(Z_1, Z_2, \ldots, Z_n \mid A, G, \Sigma) = -\frac{n}{2} \ln |\Sigma| - \frac{1}{2} \operatorname{trace} \left(\Sigma^{-1} S(A, G) \right) \tag{16}$$

where

$$S(A, G) = \sum_{t=1}^{n} X_t X_t'$$

2.6 Forecasting

The Kalman filter is the most widely used method for approximating and estimating statistics in state space models. It effectively handles changes in model parameters and variances. According to Welch and Bishop [35], during the forecasting phase, the estimated value for the current state is generated, and the error covariance is used as the initial estimate for the next step. The Kalman filter is a recursive process that begins by forming an initial state estimate, which is then updated by adding a correction to this estimate. The fundamental recursive formula is employed to update both the averages and covariance matrices [5]. Once the state space model is established, the l-step ahead forecasts from the forecast origin time t can be calculated as shown in [5].

$$\begin{aligned} \hat{Y}_t(l) &= E(Y_{t+l} \mid Y_j,\ j \le t) \qquad (3) \\ &= A\hat{Y}_t(l-1) \\ &= A \cdot A\hat{Y}_t(l-2) \\ &\vdots \\ &\vdots \\ &= A^l \hat{Y}_t. \end{aligned} \tag{17}$$

Hence,

$$\begin{aligned}\hat{Z}_t(l) &= E(Z_{t+l} \mid Z_j,\ j \le t)\\ &= H\hat{Y}_t(l)\\ &= HA^l\hat{Y}_t.\end{aligned}$$

where

$$\hat{Y}_t = E(Y_t \mid Y_j,\ j \le t) = Y_t. \tag{18}$$

As indicated in Equation (18), the accuracy of the forecasts $\hat{Z}_t(l)$ relies on the quality of the estimate $\hat{Y}_t$ of the state vector Y_t, which encapsulates the historical information necessary for future forecasting. To enhance the forecasts, new observations should be incorporated to update the state vector, thereby refining the forecast. The Kalman Filter method serves this purpose as a recursive procedure for making inferences about the state vector Y_t (see Wei [5]).

3. RESULTS AND DISCUSSION

The data used in this study are data on the IDR exchange rate against the USD, oil and gas exports, money supply and non-oil and gas exports from January 2008 to December 2019. The data source was obtained from the Ministry of Trade and Bank Indonesia website. The data plot can be seen in Figures 1-4. In Figure 1-4, it can be seen that data on the exchange rate of IDR to USD, oil and gas exports, and non-oil exports tend to experience sharp fluctuations in increase and decrease over time, causing the mean of the data in each lag to be not constant which indicates that the data is not stationary. While the money supply data tends to increase over time starting from January 2008-December 2019, so the data contains a positive trend that causes the mean of the data in each lag is not constant and the data is not stationary. The non stationary data also can be check formally by using statistical test (The ADF test), the results are given in Table 1.

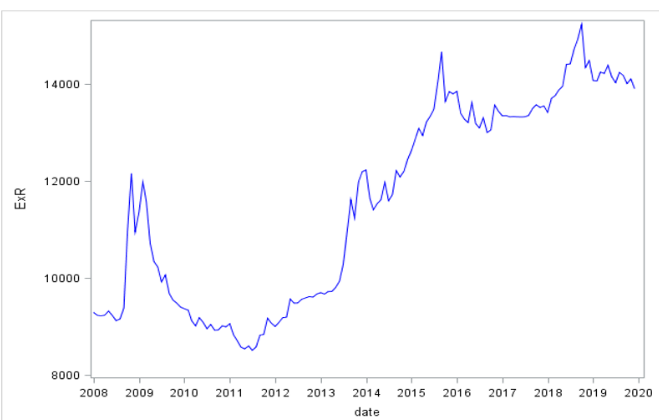

Figure 1. Plot Data IDR Exchange Rate Against the US dollars

The next step is to differentiate the data, so that the stationary in the mean can be attained. Table 2 shows the results of the ADF test for data after differentiation with $d = 1$, where the zero mean type and single mean type value with p-value <.0001 are smaller than the significant level α= 0.05 for all data variables indicating that the data are stationary. To find the best VAR(p) model, we have to determine the optimal order (lag length) p. To determine the lag length (p), we used the information criteria, which in this study uses the Akaike's Information Criterion

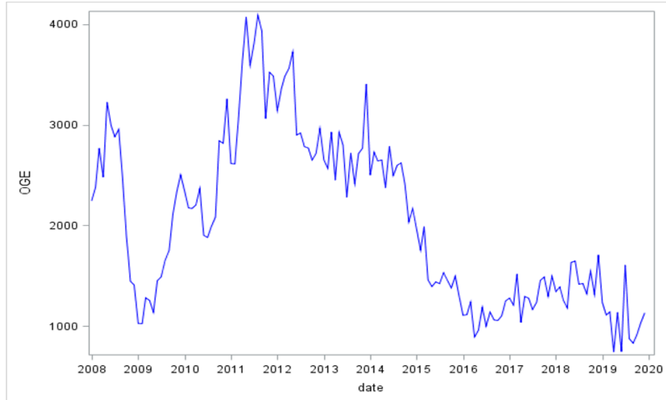

Figure 2. Plot Data Export Oil and Gas (OGE)

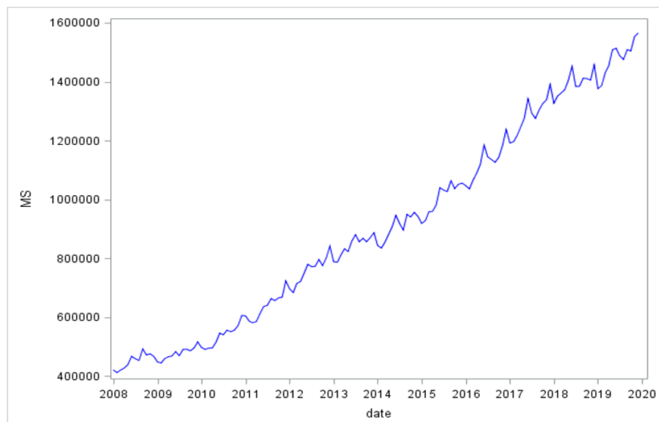

Figure 3. Plot Data Money Supply (MS)

(AIC). The optimal order of p is chosen based on the minimum value of the AIC (Table 3). From Table 3, it can be seen that the minimum AIC value is 8097.252 and found in the 6th lag. Therefore, the chosen model is VAR (6).

3.1 Model Vector Autoregressive (VAR (p))

Model VAR(6) is as follows:

$$\begin{bmatrix} Kurs_t \\ EM_t \\ UB_t \\ ENM_t \end{bmatrix} = \begin{bmatrix} -51.86032 \\ -4.50817 \\ 13627.28927 \\ 235.54144 \end{bmatrix} + \begin{bmatrix} 0.035185 & -0.14279 & 0.000159 & 0.025643 \\ -0.18916 & -0.33593 & -0.00169 & -0.00221 \\ -4.20996 & -0.17344 & -0.31181 & 2.627181 \\ 0.059396 & 0.624904 & -0.00385 & -0.90383 \end{bmatrix}$$

$$\begin{bmatrix} Kurs_{t-1} \\ EM_{t-1} \\ UB_{t-1} \\ ENM_{t-1} \end{bmatrix} + \begin{bmatrix} -0.16194 & -0.13208 & 0.002575 & 0.047211 \\ -0.06549 & -0.08526 & -0.00135 & 0.067118 \\ -10.1514 & 16.1758 & -0.16775 & 5.368599 \\ -0.19295 & 0.877631 & -0.00928 & -0.49265 \end{bmatrix} \begin{bmatrix} Kurs_{t-2} \\ EM_{t-2} \\ UB_{t-2} \\ ENM_{t-2} \end{bmatrix}$$

$$+ \begin{bmatrix} 0.085768 & -0.01582 & 0.000912 & -0.00543 \\ -0.09703 & -0.15049 & 0.000456 & 0.067269 \\ -9.99924 & -10.0912 & -0.08832 & -1.2474 \\ -0.61997 & 0.707305 & -0.00181 & -0.25497 \end{bmatrix} \begin{bmatrix} Kurs_{t-3} \\ EM_{t-3} \\ UB_{t-3} \\ ENM_{t-3} \end{bmatrix}$$

Table 3. Akaike's Information Criterion (AIC) for Model Vector Autoregressive

Information Criterion for Autoregressive Models								
Lag=0	Lag=1	Lag=2	Lag=3	Lag=4	Lag=5	Lag=6	Lag=7	Lag=8
8212.207	8145.316	8128.102	8128.199	8127.194	8110.95	8097.252	8111.454	8114.644

Table 4. Test Granger Causality

Granger-Causality Wald Test				
Test	Group Variable		p-value	Granger-cause
1	Group 1 Variables: Group 2 Variables:	ExR OGE MS non-OGE	0.0598	Non significance
2	Group 1 Variables: Group 2 Variables:	ExR OGE	0.0290	Significance
3	Group 1 Variables: Group 2 Variables:	ExR MS	0.0800	Non-significance
4	Group 1 Variables: Group 2 Variables:	ExR Non-OGE	0.5861	Non-significance
5	Group 1 Variables: Group 2 Variables:	OGE ExR MS non-OGE	0.2415	Non-significance
6	Group 1 Variables: Group 2 Variables:	OGE ExR	0.1882	Non-significance
7	Group 1 Variables: Group 2 Variables:	OGE MS	0.5476	Non-significance
8	Group 1 Variables: Group 2 Variables:	OGE Non-OGE	0.4477	Non-significance
9	Group 1 Variables: Group 2 Variables:	MS ExR OGE non-OGE	0.0119	Significance
10	Group 1 Variables: Group 2 Variables:	MS ExR	0.0787	Non-significance
11	Group 1 Variables: Group 2 Variables:	MS OGE	0.5936	Non-significance
12	Group 1 Variables: Group 2 Variables:	MS Non-OGE	0.0377	Significance
13	Group 1 Variables: Group 2 Variables:	Non-OGE ExR OGE MS	< 0.0001	Significance
14	Group 1 Variables: Group 2 Variables:	Non-OGE ExR	0.0224	Significance
15	Group 1 Variables: Group 2 Variables:	Non-OGE OGE	0.0045	Significance
16	Group 1 Variables: Group 2 Variables:	Non-OGE MS	0.0324	Significance

Notes: ExR: Exchange Rate, OGE: Oil and Gas Export, MS: Money Supply, non-OGE: Non-Oil and Gas Export

3.2 Test Granger Causality

Based on the results in Table 4, test 2 shows that the p-value = 0.0290 is smaller than the significance level of 0.05. Thus, reject the null hypothesis (Ho), meaning that future exchange rate data is influenced by past and future oil and gas export data. Furthermore, in the 9th test the decision to reject Ho is also taken because the p-value = 0.0119 is smaller than the significance level of 0.05, which means that the data of the future money supply (MS) is influenced by past and future data IDR exchange rate against the USD, oil and gas exports, money supply and non-oil and gas exports. In the 12th test, reject H0 with a p-value = 0.0377 smaller than the significance level of 0.05, so that the data of money supply (MS) will be influenced by past and future data on non-oil and gas export variables (non-OGE).

Furthermore, the 13^{th} test showed that the p-value <.0001 was smaller than the significance level of 0.05. Thus reject H0 which means data on non-oil export variables (non-OGE) that will come is influenced by past and future data on Exchange Rate, OGE, and MS. Based on the 14^{th} test the decision to reject Ho is also taken because the p-value = 0.0224 is smaller than the significance level of 0.05, which means that the data of the money supply (MS) that will come is influenced by past

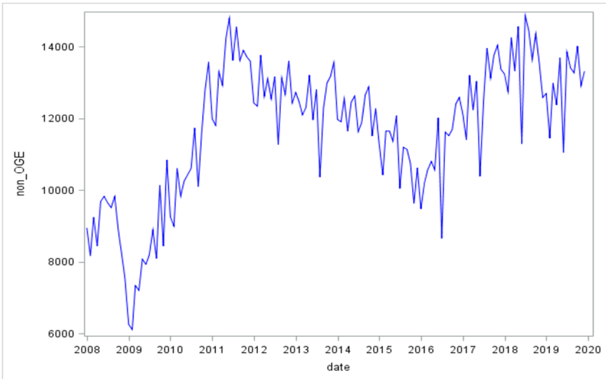

Figure 4. Plot Data Non-Oil and Gas Exports (Non-OGE)

and future data of the exchange rate variable. The 15th test shows that the p-value = 0.0045 is smaller than the significance level of 0.05, so Ho is rejected, which means the data of the future non-OGE variable is influenced by past and future data of the OGE variable. Then in the 16th test with a p-value = 0.0324 smaller than the significance level of 0.05, so reject Ho. Therefore, the future non-OGE variable data is also influenced by MS's past and future data.

As for the tests 1, 3, 4, 5, 6, 7, 8, 10, 11 with p-values are 0.0598, 0.0800, 0.5861, 0.2415, 0.1882, 0.5476, 0.4477, 0.0787, and 0.5936 respectively are higher than 0.05. Therefore, we don't reject Ho. It can be explained that, of the nine tests, all elements in group 1 in each test were affected only by themselves and not by group 2.

3.3 Canonical Correlation Analysis

We have determined the optimal lag length (p) for the Vector Autoregressive model, which is p=6, and thus selected the VAR(6) model. To construct the state space model, the first step is to choose the state vector. For this selection process, we follow the method outlined by Akaike [26], utilizing canonical correlation. The choice of state vector is based on the Information Criterion (IC) value; if the IC is negative (IC < 0), the minimum canonical correlation (ρmin) is set to zero [4]; otherwise, it is considered to be greater than zero.

From Table 5, for the first step we consider the set of state vector: $ExR_t, OGE_t, MS_t, non\text{-}OGE_t, ExR_{t+1|t}$. For this set of state vector, the value of IC is negative (−11.8368). Therefore, the variable $ExR_{t+1|t}$ is excluded from the state vector. For the second step, we consider the set of state vector: $ExR_t, OGE_t, MS_t, non\text{-}OGE_t, OGE_{t+1|t}$. For this set of state vector, the value of IC is negative (−27.8825). Therefore, the variable $non\text{-}OGE_{t+1|t}$ is excluded from the state vector. For the third step, we consider the set of state vector: $ExR_t, OGE_t, MS_t, non\text{-}OGE_t, MS_{t+1|t}$. For this set of state vector, the value of IC is positive (51.2061). Therefore, the variable $MS_{t+1|t}$ is included in the state vector. For the fourth step, we consider the set of state vector: $ExR_t, OGE_t, MS_t, non\text{-}OGE_t, MS_{t+1|t}, non\text{-}OGE_{t+1|t}$. For this set of state vector, the value of IC is positive (14.4769). Therefore, the variables $MS_{t+1|t}$ and $non\text{-}OGE_{t+1|t}$ are included in the state vector. For the fifth step, we consider the set of state vector: $ExR_t, OGE_t, MS_t, non\text{-}OGE_t, MS_{t+1|t}, non\text{-}OGE_{t+1|t}, MS_{t+2|t}$. For this set of state vector, the value of IC is positive (5.6564). Therefore, the variables $MS_{t+1|t}$, $non\text{-}OGE_{t+1|t}$, and $MS_{t+2|t}$ are included in the state vector. For the sixth step, we consider the set of state vector: $ExR_t, OGE_t, MS_t, non\text{-}OGE_t, MS_{t+1|t}, non\text{-}OGE_{t+1|t}, MS_{t+2|t}, non\text{-}OGE_{t+2|t}$. For this set of state vector, the value of IC is negative (−8.2055). Therefore, the variable $non\text{-}OGE_{t+2|t}$ is excluded from the state vector. For the seventh step, we consider the set of state vector: $ExR_t, OGE_t, MS_t, non\text{-}OGE_t, MS_{t+1|t}, non\text{-}OGE_{t+1|t}, MS_{t+2|t}, MS_{t+3|t}$.

For this set of state vector, the value of IC is negative (−8.2055). Therefore, the variable $MS_{t+3|t}$ is excluded from the state vector. Based on the analysis above, at the third, fourth, and fifth steps the IC values are positive. Therefore, based on the canonical correlation analysis, the components of the state vector are as follows:

$$Y_t = \begin{bmatrix} ExR_{t|t} \\ OGE_{t|t} \\ MS_{t|t} \\ \text{non-}OGE_{t|t} \\ MS_{t+1|t} \\ \text{non-}OGE_{t+1|t} \\ MS_{t+2|t} \end{bmatrix}$$

3.4 State Space Model

Based on the estimation of parameters given in Table 6, then model state space can be written as follows:

$$Y_{t+1} = AY_t + GX_{t+1} \tag{19}$$

$$\begin{bmatrix} ExR_{t+1|t+1} \\ OGE_{t+1|t+1} \\ MS_{t+1|t+1} \\ \text{non-}OGE_{t+1|t+1} \\ MS_{t+2|t+1} \\ \text{non-}OGE_{t+2|t+1} \\ MS_{t+3|t+1} \end{bmatrix} = \begin{bmatrix} 0.0333 & -0.0864 & -0.0016 & -0.0093 & 0 & 0 & 0 \\ -0.1421 & -0.2832 & -0.0009 & -0.0205 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ -0.4832 & -0.6218 & -0.0058 & 0.5610 & -0.00439 & 0.31402 & -0.0252 \\ -18.145 & -31.0828 & 0.18733 & 35.9854 & -0.03575 & 57.2744 & -0.8375 \end{bmatrix} \begin{bmatrix} ExR_{t|t} \\ OGE_{t|t} \\ MS_{t|t} \\ \text{non-}OGE_{t|t} \\ MS_{t+1|t} \\ \text{non-}OGE_{t+1|t} \\ MS_{t+2|t} \end{bmatrix} + \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0.2427 & -1.11634 & -0.4058 & 5.9140 \\ 0.1644 & 0.4149 & -0.0034 & -0.6956 \\ -6.1543 & -13.6988 & -0.0563 & 0.0307 \end{bmatrix} \begin{bmatrix} \alpha_{t+1} \\ \beta_{t+1} \\ \gamma_{t+1} \\ \delta_{t+1} \end{bmatrix} \tag{20}$$

where

$$\text{Var}\begin{bmatrix} \alpha_{t+1} \\ \beta_{t+1} \\ \gamma_{t+1} \\ \delta_{t+1} \end{bmatrix} = \begin{bmatrix} 106314.9 & -10423.2 & 1782192 & -38498.4 \\ -10423.2 & 84627.4 & 1009557 & 90274.95 \\ 1782192 & 1009557 & 4.2142 \times 10^8 & 3036868 \\ -38498.4 & 90274.95 & 3036868 & 828225.2 \end{bmatrix} \tag{21}$$

3.5 Forecasting

In this study, forecasting is carried out using a fitted state space model and the Kalman filter technique. The state space model is applied to predict data for the upcoming 12 periods (months). From the results of

Table 5. Canonical Correlation Analysis

STATE VECTOR	Canonical Correlation	IC	Chi-Square	DF
ExR_t, OGE_t, MS_t, non-OGE_t, $ExR_{t+1\|t}$	1, 1, 1, 1, 0.4727	-11.8368	33.1285	24
ExR_t, OGE_t, MS_t, non-OGE_t, $OGE_{t+1\|t}$	1, 1, 1, 1, 0.3622	-27.8825	18.4293	24
ExR_t, OGE_t, MS_t, non-OGE_t, $MS_{t+1\|t}$	1, 1, 1, 1, 0.7073	51.2061	90.8811	24
ExR_t, OGE_t, MS_t, non-OGE_t, $MS_{t+1\|t}$, non-$OGE_{t+1\|t}$	1, 1, 1, 1, 0.7178, 0.5872	14.4769	55.6134	23
ExR_t, OGE_t, MS_t, non-OGE_t, $MS_{t+1\|t}$, non-$OGE_{t+1\|t}$, $MS_{t+2\|t}$	1, 1, 1, 1, 0.7194, 0.6942, 0.5416	5.6564	45.8366	22
ExR_t, OGE_t, MS_t, non-OGE_t, $MS_{t+1\|t}$, non-$OGE_{t+1\|t}$, $MS_{t+2\|t}$, non-$OGE_{t+2\|t}$	1, 1, 1, 1, 0.7195, 0.6944, 0.5956, 0.4587	-8.2055	31.3130	21
ExR_t, OGE_t, MS_t, non-OGE_t, $MS_{t+1\|t}$, non-$OGE_{t+1\|t}$, $MS_{t+2\|t}$, $MS_{t+3\|t}$	1, 1, 1, 1, 0.7747, 0.6960, 0.6606, 0.4884	-3.0178	36.1198	21

Notes: ExR: Exchange Rate, OGE: Oil and Gas Export, MS: Money Supply, non-OGE: Non-Oil and Gas Export, IC: Information Criterion.

Table 6. Parameters Estimate of the State Space Model

Parameter Estimates							
Parameter	Estimate	Standard Error	t Value	Parameter	Estimate	Standard Error	t Value
F(1,1)	0.033493	0.081146	0.41	F(7,3)	0.187334	0.190972	0.98
F(1,2)	-0.08594	0.089877	-0.96	F(7,4)	35.98542	2.236108	16.09
F(1,3)	-0.00161	0.001068	-1.51	F(7,5)	-0.03575	0.267784	-0.13
F(1,4)	-0.00891	0.023787	-0.37	F(7,6)	57.27446	2.478451	23.11
F(2,1)	-0.14186	0.068716	-2.06	F(7,7)	-0.83752	0.301143	-2.78
F(2,2)	-0.28272	0.076873	-3.68	G(5,1)	0.242711	2.366285	0.10
F(2,3)	-0.00096	0.000947	-1.01	G(5,2)	-1.11634	2.405108	-0.46
F(2,4)	-0.02041	0.020514	-1.00	G(5,3)	-0.40586	0.082366	-4.93
F(6,1)	-0.48324	0.133432	-3.62	G(5,4)	5.914061	1.546771	3.82
F(6,2)	-0.62181	0.229897	-2.70	G(6,1)	0.164405	0.090163	1.82
F(6,3)	-0.00583	0.003292	-1.77	G(6,2)	0.414905	0.115255	3.60
F(6,4)	0.561010	0.201008	2.79	G(6,3)	-0.00341	0.003183	-1.07
F(6,5)	-0.00439	0.005288	-0.83	G(6,4)	-0.69560	0.052994	-13.13
F(6,6)	0.314026	0.303280	1.04	G(7,1)	-6.15438	2.358714	-2.61
F(6,7)	-0.02523	0.010446	-2.42	G(7,2)	-13.6988	2.421685	-5.66
F(7,1)	-18.1456	2.592113	-7.00	G(7,3)	-0.05636	0.083116	-0.68
F(7,2)	-31.0828	2.602168	-11.94	G(7,4)	0.030778	1.574403	0.02

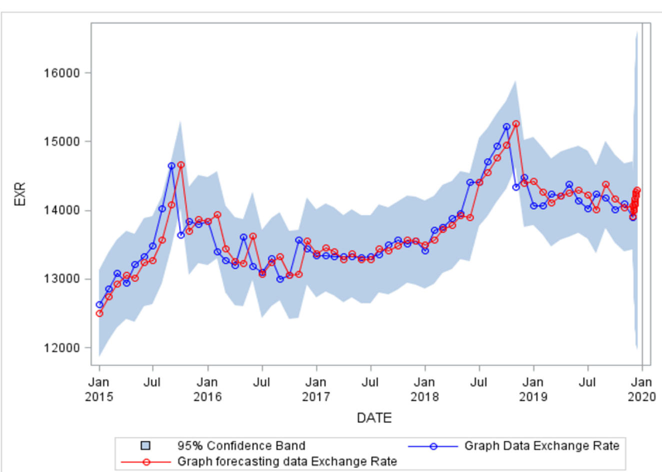

Figure 5. Plot Forecasting for Data Exchange Rate

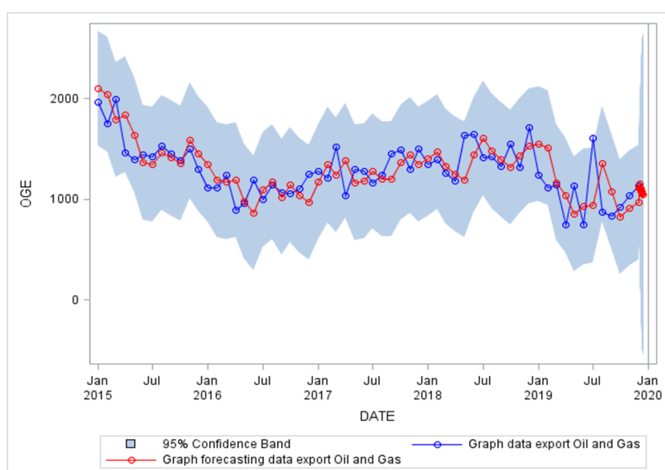

Figure 6. Plot Data Forecasting Data Export Oil and Gas

Table 7. Forecasting Data for the Next 12 Months

Month	ExR	OGE	MS	Non-OGE
January 2020	13906.54	1115.56	1538891.46	13229.69
February 2020	13995.18	1149.06	1551906.00	13483.11
March 2020	14015.69	1112.25	1565624.31	13571.25
April 2020	14040.30	1107.69	1571544.97	13690.38
May 2020	14074.59	1100.27	1591046.97	13490.64
June 2020	14090.36	1085.76	1598833.64	13467.48
July 2020	14123.46	1083.52	1596261.97	13551.67
August 2020	14171.85	1083.10	1603880.41	13638.64
September 2020	14204.12	1070.16	1614198.62	13718.18
October 2020	14232.63	1060.61	1623277.94	13748.45
November 2020	14263.17	1052.83	1634547.66	13709.10
December 2020	14290.71	1043.59	1642125.15	13710.71

Notes: ExR: Exchange Rate, OGE: Oil and Gas Export, MS: Money Supply, Non-OGE: Non-Oil and Gas Export.

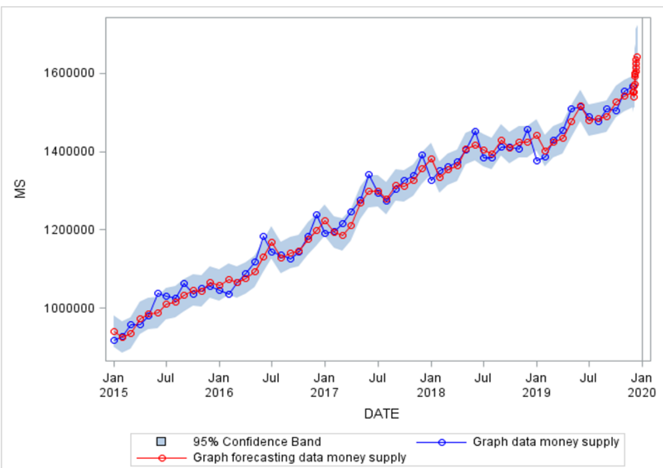

Figure 7. Plot Forecasting for Data Money Supply

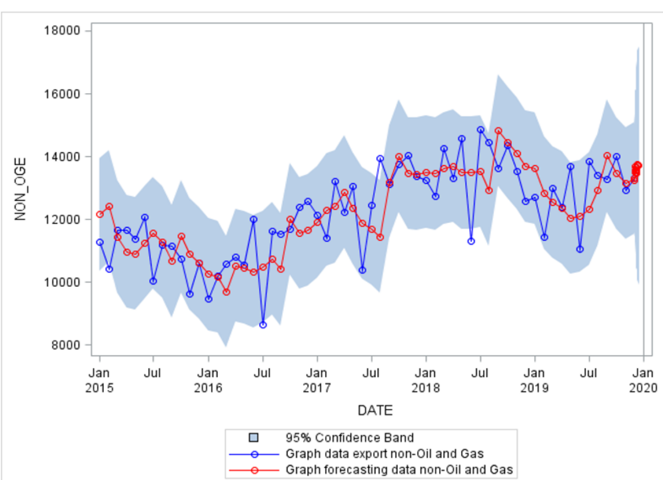

Figure 8. Plot Forecasting Data Export Non-Oil and Gas

forecasting given in Table 7, for data IDR exchange rate with respect to US dollars (ExR), the trend of the forecast for the next 12 months is increase (Table 7 and Figure 5), for data IDR exchange rate with respect to US dollars (ExR), the trend of the forecast for the next 12 months is increase (Table 7 and Figure 5), for data Export of Oil and Gas (OGE), the trend of the forecast for the next 12 months is decrease (Table 7 and Figure 6), for data money supply (MS), the trend of the forecast for the next 12 months is increase (Table 7 and Figure 7), for data expoert of non-Oil and Gas (non-OGE), the trend of the forecast for the next 12 months is increase, but very small increment (Table 7 and Figure 8). The predicted values, the real and 95% confidence interval for the data on the Indonesia Rupiah (IDR) exchange rate (ExR) against the US Dollar (USD), oil and gas exports (OGE), money supply (MS) and non-oil and gas exports (non-OGE) from January 2008 to December 2019 are given in Figures 5-8.

4. CONCLUSIONS

Based on the results of the study on the Application of State Space Representation in the Vector Autoregressive (VAR) Model for Multivariate Time Series Data Forecasting on IDR Exchange Rates against USD (ExR), Oil and Gas Exports (OGE), Money Supply (MS), and Non-Oil and Gas Exports (non-OGE). This study focuses on determining the best model for modeling the multivariate time series data ExR, OGE, MS and non-OGE by using VAR(p) model then the state space model for forecasting. Based on the results of analysis and the minimum value of Akaike's Information Criterion (AIC), the best VAR model is VAR (6) model. Based on this VAR (6) model, the state space model is derived by using procedure of canonical correlation to find the state vector. Furthermore, the behavior of data also discussed by using granger causality; and forecasting is conducted by using state space model for the next 12 months. The results of forecasting values by using state space model are sound, and this is indicated by the predicted values and the real values for the state space model are very closed to each other.

5. ACKNOWLEDGEMENT

The authors thank the Ministry of Trade and Bank Indonesia for providing data in this study.

REFERENCES

[1] G. Box and G. Jenkins. *Time Series Analysis: Forecasting and Control*. Holden-Day, San Francisco, 1976.

[2] P. J. Brockwell and R. A. Davis. *Time Series: Theory and Methods*. Springer Verlag, New York, 2nd edition, 1991.

[3] P. J. Brockwell and R. A. Davis. *Introduction to Time Series Analysis and Forecasting*. Springer Verlag, New York, 2nd edition, 2002.

[4] R. S. Tsay. *Analysis of Financial Time Series*. John Wiley & Sons, Inc., New York, 2005.

[5] W. W. S. Wei. *Time Series Analysis: Univariate and Multivariate Methods*. Pearson Education, New York, 2nd edition, 2006.

[6] J. D. Hamilton. *Time Series Analysis*. Princeton University Press, New Jersey, 1994.

[7] R. S. Tsay. *Multivariate Time Series Analysis*. John Wiley & Sons, Inc., New York, 2014.

[8] J. Durbin and S. J. Koopman. *Time Series Analysis by State Space Methods*. Oxford University Press, Oxford, 2nd edition, 2012.

[9] R. E. Kalman. A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, 82(1):35–45, 1960.

[10] R. E. Kalman and R. S. Bucy. New results in linear filtering and prediction theory. *Journal of Basic Engineering*, 83(3):95–108, 1961.

[11] M. H. A. Davis and R. B. Vinter. *Stochastic Modelling and Control*. Chapman and Hall, London, 1985.

[12] E. J. Hannan and M. Deistler. *The Statistical Theory of Linear Systems*. John Wiley and Sons, Inc., New York, 1988.

[13] H. Akaike. Information theory and an extension of the maximum likelihood principle. In B. N. Petrov and F. Csaki, editors, *Second International Symposium on Information Theory*, pages 267–281, Budapest, 1973. Akademiai Kiado.

[14] H. Akaike. Markovian representation of stochastic processes and its application to the analysis of autoregressive moving average processes. *Annals of the Institute of Statistical Mathematics*, 26:363–387, 1974.

[15] A. H. Jazwinski. *Stochastic Processes and Filtering Theory*. Academic Press, New York, 1970.

[16] B. D. O. Anderson and J. B. Moore. *Optimal Filtering*. Prentice-Hall, Englewood Cliffs, 1979.

[17] M. Aoki and A. Havenner. State space modelling of multiple time series. *Econometric Reviews*, 10:1–59, 1991.

[18] A. C. Harvey. *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge University Press, Cambridge, 1989.

[19] A. C. Harvey. *Time Series Models*. Harvester Wheatsheaf, Hemel Hempstead, 2nd edition, 1993.

[20] M. West and J. Harrison. *Bayesian Forecasting and Dynamic Models*. Springer-Verlag, New York, 2nd edition, 1997.

[21] C.-J. Kim and C. R. Nelson. *State-Space Models with Regime Switching, Classical and Gibbs-Sampling Approaches with Applications*. MIT Press, Cambridge, MA, 1999.

[22] R. H. Shumway and D. S. Stoffer. *Time Series Analysis and Its Applications*. Springer-Verlag, New York, NY, 2000.

[23] N. H. Chan. *Time Series: Application to Finance*. John Wiley and Sons, Inc., New York, 2002.

[24] V. Gomez. *Multivariate Time Series with Linear State Space Structure*. Springer Verlag, Switzerland, 2016.

[25] H. Akaike. Markovian representation of stochastic processes by canonical variables. *SIAM Journal on Control*, 13:162–173, 1975.

[26] H. Akaike. Canonical correlation analysis of time series and the use of an information criterion. In R. Mehra and K. Lainiotis, editors, *System Identification: Advances and Case Studies*. Academic Press, Inc., New York, 1976.

[27] J. Durbin. Introduction to state space time series analysis. In Andrew Harvey, S. J. Koopman, and Neil Shephard, editors, *State Space and Unobserved Component Models Theory and Applications*. Cambridge University Press, London, 2004.

[28] E. Virginia, J. Ginting, and F. A. M. Elfaki. Application of garch model to forecast data and volatility of share price of energy (study on adaro energy tbk, lq45). *International Journal of Energy Economics and Policy*, 8(3):131–140, 2018.

[29] Warsono, E. Russel, Wamiliana, Widiarti, and M. Usman. Vector autoregressive with exogenous variable model and its application in modeling and forecasting energy data: Case study of ptba and hrum energy. *International Journal of Energy Economics and Policy*, 9(2):390–398, 2019.

[30] Warsono, E. Russel, Wamiliana, Widiarti, and M. Usman. Modeling and forecasting by the vector autoregressive moving average model for export of coal and oil data. *International Journal of Energy Economics and Policy*, 9(4):240–247, 2019.

[31] L. M. Hamzah, S. U. Nabila, E. Russel, M. Usman, E. Virginia, and Wamiliana. Dynamic modeling and forecasting of data export of agriculture commodity by vector autoregressive model. *Journal of Southwest Jiaotong University*, 55(3):1–10, 2020.

[32] W. Enders. *Applied Econometric Time Series*. John Wiley and Sons, New York, 4th edition, 2015.

[33] Warsono, E. Russel, A. Putri, Wamiliana, Widiarti, and M. Usman. Dynamic modeling using vector error-correction model studying the relationship among data share price of energy pgas malaysia, akra, indonesia, and ptt pcl-thailand. *International Journal of Energy Economics and Policy*, 10(2):360–373, 2020.

[34] H. Lüthkepohl. *New Introduction to Multiple Time Series Analysis*. Springer-Verlag, Berlin, 2005.

[35] G. Welch and G. Bishop. An introduction to the kalman filter. Technical report, 2001.