BAREKENG: Journal of Mathematics and Its Application
December 2022 Volume 16 Issue 4 Page 1313–1324
P-ISSN: 1978-7227 E-ISSN: 2615-3017
https://doi.org/10.30598/barekengvol16iss4pp1313-1324

THE PROMINENCE OF VECTOR AUTOREGRESSIVE MODEL IN MULTIVARIATE TIME SERIES FORECASTING MODELS WITH STATIONARY PROBLEMS

Embay Rohaeti [1], I Made Sumertajaya[2*], Aji Hamim Wigena[3], Kusman Sadik[4]

[1,2,3,4]Departement of Statistics, Faculty of Mathematics and Natural Sciences, IPB University
Jl. Meranti IPB University, Bogor, 16680, West Java, Indonesia
[1]Departement of Mathematics, Faculty of Mathematics and Natural Sciences, Pakuan University
Jl. Pakuan PO.BOX 452, Bogor,16143, West Java, Indonesia

Corresponding author's e-mail: [2] imsjaya@apps.ipb.ac.id*

Abstract. *One of the problems in modeling multivariate time series is stationary. Stationary test results do not always produce all stationary variables; mixed stationary and non-stationary variables are possible. When stationary problems are found in multivariate time series modeling, it is necessary to evaluate the model's performance in various stationary conditions to obtain the best forecasting model. This study aims to get a superior multivariate time series forecasting model based on the goodness of the model in various stationary conditions. In this study, the evaluation of the model's performance through simulation data modeling is then applied to the actual data with a stationary problem, namely Bogor City inflation data. The best model in simulation modeling is based on the stability of RMSE and MAD in 100 replications. The results are that the VAR model is the best in various stationary conditions. Meanwhile, the best model on actual data modeling is based on evaluation in 4 folds for model fitting power and model forecasting power. The Bogor City inflation data modeling with the mixed stationary problem resulted in the best model, namely the VAR(1) model. This means the VAR model is good enough to be used as a forecasting model in mixed stationary conditions. Thus, in this study, based on the goodness of the model in two modeling scenarios in various stationary conditions, overall, it was found that the VAR model was superior to the VARD and VECM models.*

Keywords: *Stationary, Simulation, VAR, VARD, VECM, Inflation.*

Article info:

Submitted: 6th July 2022 *Accepted: 16th October 2022*

How to cite this article:

E. Rohaeti, I. M. Sumertajaya, A. H. Wigena and K. Sadik, "THE PROMINENCE OF VECTOR AUTOREGRESSIVE MODEL IN MULTIVARIATE TIME SERIES FORECASTING MODELS WITH STATIONARY PROBLEMS", *BAREKENG: J. Math. & App.*, vol. 16, iss. 4, pp. 1313-1324, Dec., 2022.

1. INTRODUCTION

In modeling multivariate time series data, there are several model choices, including the vector autoregressive (VAR) model, vector autoregressive with differencing (VARD), and vector error correction model (VECM). Stationarity can be a problem in time series modeling, so testing is necessary. In addition to visual, stationary testing can be carried out with formal tests, including the augmented Dickey-Fuller (ADF) test [1]. The results of the ADF test do not always result in all variables being stationary as expected. Another possibility is to produce a mixed stationary variable, or all variables are not stationary.

Multivariate time series data are increasingly available in various fields, for example, in the economic area, namely inflation data. When economic conditions are unstable, one of the indicators is inflation. This phenomenon causes problems that can affect regional and national economic growth [2]. Inflation is an essential focus of the regional (TPID) and national (TPIN) inflation control teams, so the inflation rate needs to be controlled. One of the efforts to control inflation is to predict future inflation. This means that a good forecasting model is required to control inflation.

Previous studies include [3] analyzing the cointegration and causality of Engel Granger deposit interest rates, changes in the IDR/USD exchange rate, and stock market returns on the Jakarta Stock Exchange. [4] analyzed the autoregressive vector (VAR) on the interrelationship between the financing deposit ratio (FDR) and returned on assets (ROA) in Islamic banks in Indonesia. [5] analyzed the influence of the supplier region area on the price of red chilli in DKI Jakarta using VECM. [6] modeled and predicted export data for agricultural commodities using the VAR model. [7] A stationary test of the Padang city inflation data for 2014-2019. [8] driven multivariate modeling of tourist visits to Indonesia through air, sea, and land gates involving the impact of the covid 19 outbreak.

The difference between this research and previous research is that previous research has not done modeling with simulation data. Besides, the actual data modeling has not been evaluated in various scenarios of multivariate time series models and different stationary conditions. Based on these differences, modeling with simulation data was carried out in this study, and the modeling was evaluated in various stationary conditions and scenarios of multivariate time series models. The simulation data modeling was assessed on 100 replicates in various stationary conditions and multivariate time series model scenarios. It aims to obtain stability of model performance in different multivariate time series modeling scenarios. In addition, the actual data modeling applies the K-fold concept to get the best model in various conditions from the same data [9]. Furthermore, it is used for the actual data containing the stationary problem, namely the case of Bogor City inflation data. The accuracy of the model performance evaluation results is based on the smallest RMSE value for the evaluation of model fitting power and forecasting power. This study aims to obtain a multivariate time series forecasting model that excels in various stationary conditions or when there is a stationary problem in multivariate time series data. So that the advantage of this study is that a multivariate time series forecasting model is obtained that is superior in various stationary conditions.

2. RESEARCH METHODS

This study uses simulation data and actual data. Actual data obtained from the Central Bureau of Statistics Indonesia (BPS) is multivariate time series data in the form of monthly inflation data in the foodstuff group from January 2014 to December 2019. The variables used in the modeling are limited to three variables, namely preserved fish (y_1), fresh fish (y_2) and vegetables (y_3). Modeling simulation data in 100 replications evaluated the multivariate time series model. It aims to obtain the stability of the best model performance in three modeling scenarios in various stationary conditions. Furthermore, it is applied to the actual data containing the stationary problem, namely the case of inflation data for the Bogor City food ingredients group. The accuracy of the model performance evaluation results is based on the smallest RMSE value for the evaluation of model fitting power and forecasting power.

2.1 Modeling Using Generated Data or Simulation Data

The stages of analysis in modeling with simulation data are

a. The first stage is to generate simulation data.

b. Simulation data is generated by following the VAR(1) model with three variables. The length of the observation period (t) of the generated data is 240-time points.

c. The scenario used in the simulation is that all variables are stationary, the variables are mixed stationary, and all variables are not stationary.

d. Tested for stationery using Augmented Dickey-Fuller (ADF) test.

e. Performed multivariate time series modeling for each scenario.

f. Evaluate the three models by calculating the RMSE and MAD values

g. The process is repeated 100 times to obtain the best model.

2.2 Modeling with Actual Data

The stages of analysis in modeling with actual data are

a. Data exploration

   This stage aims to see the data pattern from all variables or to present an overview of the data [7], [10].

b. Stationary test

   The hypothesis used in the ADF test is

   H_0 : data contains unit root or is not stationary $(A = 0)$

   H_1 : data does not contain unit root or is stationary $(A < 0)$

   The test statistic:

$$t_{test} = \frac{\hat{A}}{\sigma_{\hat{A}}} \tag{1}$$

   where A is the intercept, $\hat{A}$ is the estimated value of A, and σ is the standard deviation of $\hat{A}$. The criteria for rejecting H_0 are if the p-value < $\alpha = 5\%$. It shows that the observed time series data does not contain unit roots or stationary data. Otherwise, the decision accepts H_0. It shows that the observed time series data contains unit roots or the data is not stationary [6], [10], [11]. When non-stationary results are obtained, differencing is performed [10], [11].

   According to [10] and [12], cointegration occurs when the movement pattern is the same between periods in a long period on a time series variable. The cointegration test used is the Johansen cointegration test with trace test type. The test's null hypothesis (H_0) is that there is no cointegration equation between variables. The alternative hypothesis (H_1) states that at least one cointegration relationship is formed between time series variables.

c. Determination of optimum lag

   [12] generally presents the criteria used to determine the lag length of the VAR model, namely the Akaike criteria (AIC), Schwarz criteria (SC), Hannan-Quinn criteria (HQ), and the final prediction error (FPE) criteria. The following are the criteria for determining the length of the lag in multivariate time series:

$$AIC(\rho) = \ln|\Sigma(\rho)| + \frac{2}{T}\rho s^2 \tag{2}$$

$$SC(\rho) = \ln|\Sigma(\rho))| + \frac{\ln T}{T}\rho s^2 \tag{3}$$

$$HQ(\rho) = \ln|\Sigma(\rho))| + \frac{2\ln(\ln T)}{T}\rho s^2 \tag{4}$$

$$FPE(\rho) = \left[\frac{T+s\rho+1}{T-s\rho-1}\right]^k |\Sigma(\rho)| \tag{5}$$

where $\Sigma(\rho)$ is the covariance matrix, s is the number of variables, T is the number of observations, and p is the order. The optimal lag length is obtained when the criteria are used to produce the smallest value [13], [14]. The lag selection in the VAR model is also known as the order selection of the VAR model.

d. Multivariate time series modeling (VAR, VARD, VECM)

According to [1], [10], [15], [16], [17], [18] the VAR model is a system of regression equations in which each variable is regressed against other variables, including the variable itself at the previous time point. The general form of the VAR (p) model is as follows:

$$\boldsymbol{y}_t = \boldsymbol{A}_0 + \boldsymbol{A}_1\boldsymbol{y}_{t-1} + \boldsymbol{A}_2\boldsymbol{y}_{t-2} + \cdots + \boldsymbol{A}_p\boldsymbol{y}_{t-p} + \boldsymbol{u}_t,\ \ t = 1, \dots, T \tag{6}$$

Equation (6) can be written in the form:

$$\boldsymbol{y}_t = \boldsymbol{A}_0 + \sum_{i=1}^{p} \boldsymbol{A}_i\boldsymbol{y}_{t-i} + \boldsymbol{u}_t \tag{7}$$

with $\boldsymbol{y}_t, \boldsymbol{y}_{t-i}$ is a vector of size $(n \times 1)$. It contains n variables included in the VAR model at times t and $t-i$, where $i = 1,2, \dots p,$ and p is the order of VAR, and t is the observation period. $\boldsymbol{A}_0 = (A_{10}, A_{20}, \dots, A_{n0})'$ is an intercept vector of size $(n \times 1)$. $\boldsymbol{A}_i$ is a coefficient matrix of size $(n \times n)$, and $\boldsymbol{u}_t = (u_{1t}, u_{2t}, \dots u_{nt})'$ is a white noise error vector of size (n×1).

The VAR (p) model is developed into a vector error correction model (VECM) when the data is not stationary and has one or more cointegration relationships. [1], [6], [10], [16] present the general form of VECM as follows:

$$\Delta\boldsymbol{y}_t = \boldsymbol{A}_0 + \boldsymbol{\pi}\boldsymbol{y}_{t-1} + \sum_{i=1}^{p-1} \boldsymbol{\Gamma}_i\Delta\boldsymbol{y}_{t-i} + \boldsymbol{u}_t \tag{8}$$

with Δ is the difference operator where $\Delta\boldsymbol{y}_t = \boldsymbol{y}_t - \boldsymbol{y}_{t-1}$, and $\boldsymbol{y}_{t-1}$ is a time series variable vector in the first lag of size $(n \times 1)$. $\boldsymbol{A}_0$ is an intercept of size $(n \times 1)$, $\boldsymbol{\pi}$ is a cointegration coefficient matrix of size $(n \times k)$, $\boldsymbol{\Gamma}_i$ is a coefficient of the ith-variable matrix of size $(n \times n)$, with $i = 1,2, \dots p-1$, and $\boldsymbol{u}_t$ is an error vector of size $(n \times 1)$.

e. Evaluation of the goodness of model

According to [18] and [19] , the Root Mean Squared Error (RMSE) can measure the goodness of the model. RMSE is the amount of deviation from the predicted value to the actual value. RMSE is formulated as follows:

$$RMSE = \left(\frac{\sum_{t=1}^{n}(y_t - \hat{y}_t)}{n}\right)^{1/2} \tag{9}$$

where y_t is the observed value at time t, and $\hat{y}_t$ is the estimated value at time t. Other measures, such as the Akaike Information Criterion (AIC) and Mean Absolute Deviation (MAD). The goodness of the model can be determined by the smallest value of RMSE, AIC, or MAD [18], [20].

f. Impulse response function (IRF) and forecast error variance decomposition (FEVD) testing [5], [17].

g. Diagnostic test on the best model [20].

h. Forecasting using the best model

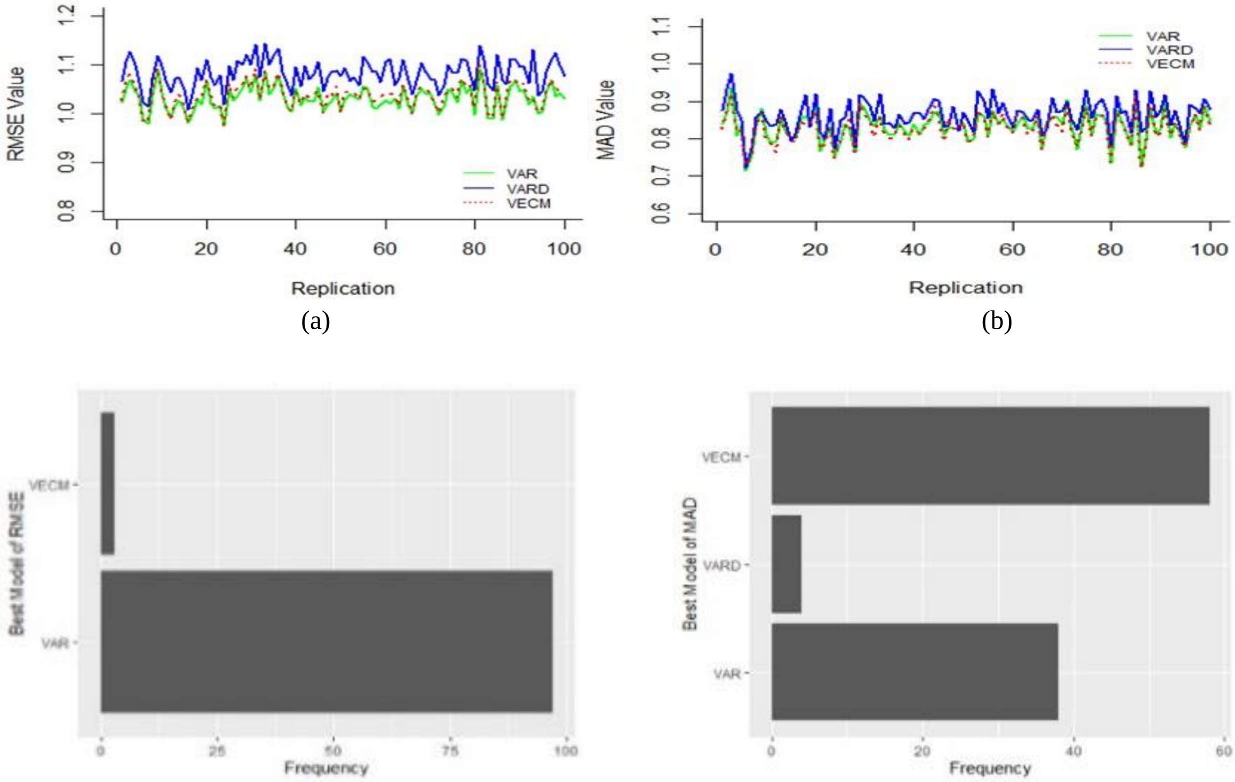
3. RESULTS AND DISCUSSION

3.1 Modeling on Simulation Data

The study of the multivariate time series model on the simulation data starts with generating three variables y_t which will be modeled following the VAR(1) model. Previously, generated random numbers following the standard normal distribution $u_t \sim N(0,1)$ of 300 data for each variable. Then determine the initial value of the data at time $t = 1$. Furthermore, the data burning process is carried out by removing the first 60 data of y_t on each variable so that the length of the observation period (t) of simulation data is 240-time points. Furthermore, the stationarity test was carried out on the variables. Finally, modeling is carried out with three scenarios.

3.1.1 Scenario 1: All variables are not stationary

The first scenario in this study is modeling with all non-stationary variables. This modeling begins with the generation of simulation data on the condition that all variables (y_1, y_2, y_3) are not stationary. Next are VAR, VARD, and VECM modeling. The performance of the three models was evaluated based on the stability of the RMSE and MAD in 100 replicates. The results of the performance evaluation of the three models in 100 replications are presented in Figure 1.



(a) (b)

(c) (d)

Figure 1. Evaluation of the Three Models when All Variables are not Stationary (a) RMSE Stability on Three Models, (b) MAD Stability on Three Models, (c) Best Model Based on RMSE, (d) Best Model Based on MAD

Figures 1(a) and 1(b) show the stability of the performance of the three models based on the evaluation of RMSE and MAD in 100 replications. The results show that the VAR and VECM models perform almost equally well. Figures 1(c) and 1(d) present the best model in 100 replications. The results obtained by the VAR model as the best model based on the RMSE evaluation. This is obtained when all non-stationary variables are different. Meanwhile, based on the MAD evaluation, the best model was obtained, namely, the VECM model, when no differencing was performed on all variables. So, in this scenario simulation, it can be said that the VAR model has almost as good a performance as VECM. It means that the VAR model is still suitable for modeling with the condition that all variables are not stationary.

3.1.2 Scenario 2: Mixed stationary variable

The second scenario is modeling with mixed stationary variables. This scenario simulation modeling is the same as the first scenario simulation modeling process, but y_1, y_2, and $y_3$are generated with mixed stationary conditions. The results of the evaluation of the performance of the VAR, VARD, and VECM models in 100 replications under mixed stationary conditions are presented in Figure 2.

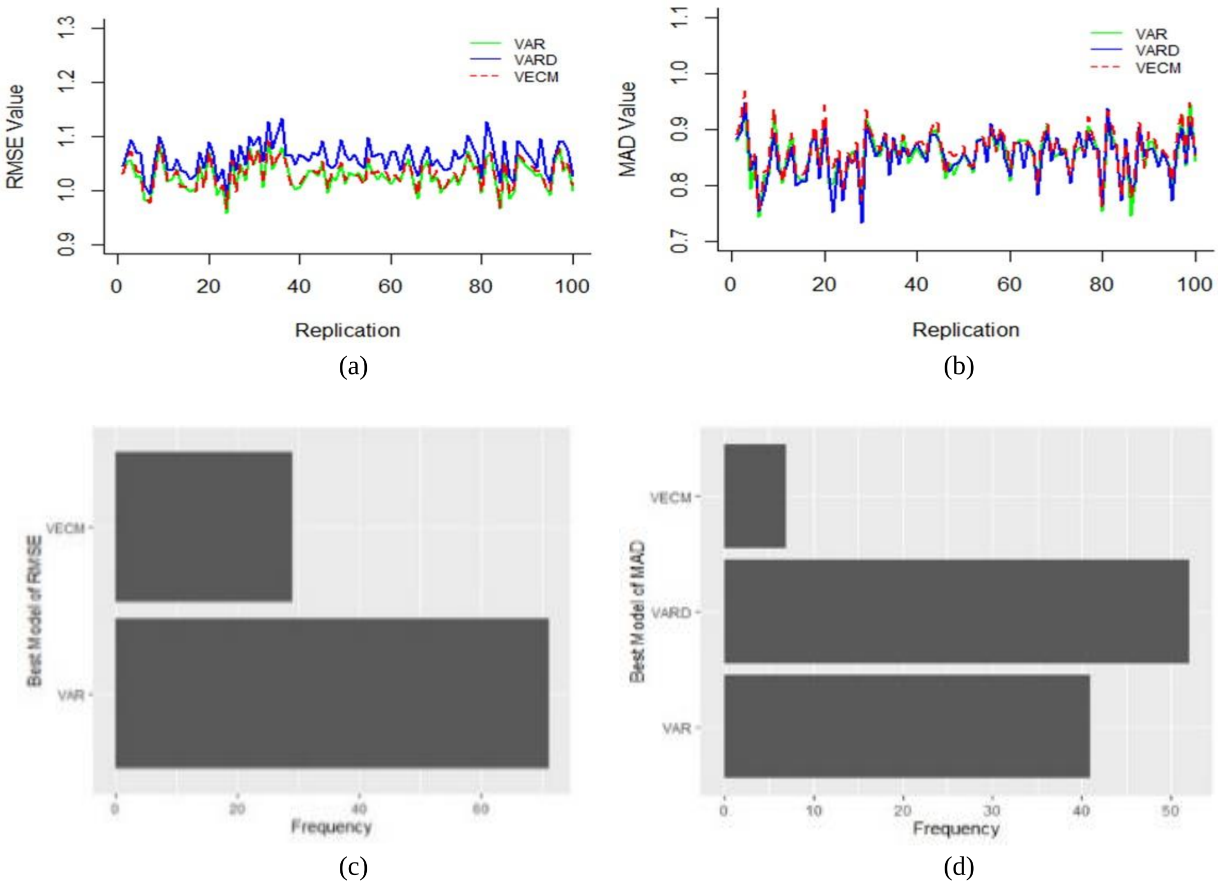


Figure 2. Model Evaluation in Mixed Stationary Scenario, (a) RMSE Stability on Three Models, (b) MAD Stability on Three Models, (c) Best Model Based on RMSE, (d) Best Model Based on MAD

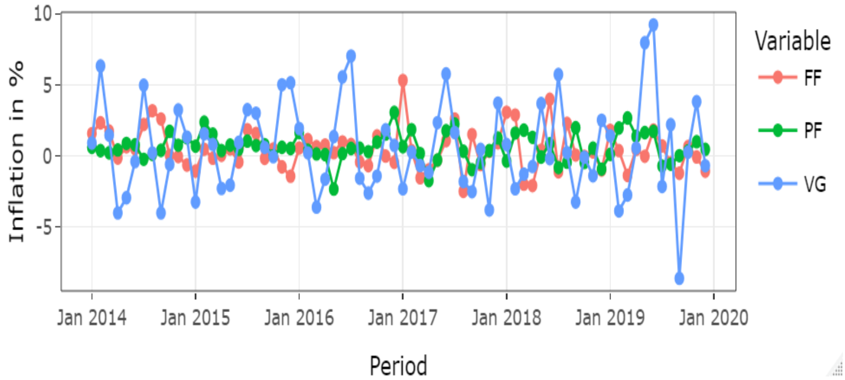
Figures 2(a) and 2(b) show the stability of the three models based on the evaluation of RMSE and MAD in 100 replicates. The results show that the RMSE in the VAR and VECM models are almost the same, while the MAD values for the three models are almost the same. The best model in 100 replications is shown in Figures 2(c) and 2(d). The results obtained by the VAR model as the best model based on the RMSE evaluation. Meanwhile, based on the MAD evaluation, the best model was obtained, namely the VARD model. These results were obtained when the non-stationary variables were differencing at order 1. This means that in this scenario simulation, it can be said that the VAR model has almost as good a performance as the VARD. Thus, the VAR model is still well used in modeling with mixed stationary conditions.

3.2 Modeling on Secondary Data (Actual Data)

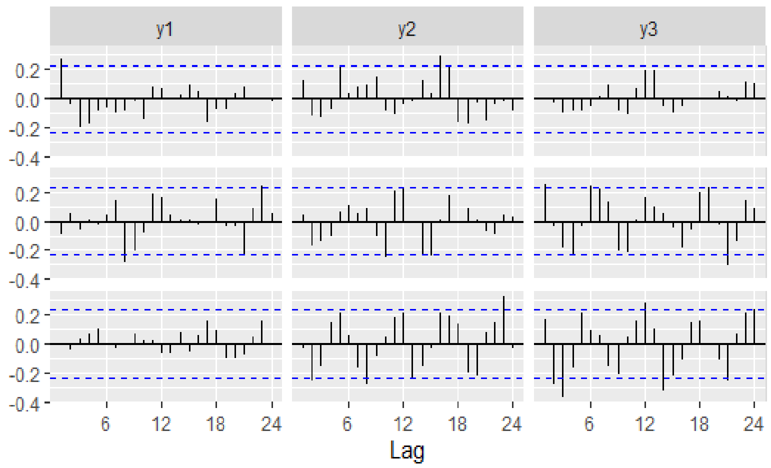
The actual data used is the inflation data for the Bogor City foodstuff group from January 2014 to December 2019. The variables used are limited to three variables, namely preserved fish (y_1), fresh fish (y_2), and vegetables (y_3) . The stages in this modeling begin with actual data exploration, stationary testing, dividing the data 4-fold, determining the optimal lag, modeling with three scenarios, diagnostic test for the best model, testing IRF and FEVD, and forecasting.

3.2.1 Data Exploration

Stationarity can be checked visually through data exploration. The results of the exploration of the three variables show that the variable y_1 is not stationary, while the variables y_2 and y_3 are already stationary. The results of data exploration are presented in Figure 3 below:



(a)



(b)

Figure 3. Exploration on Bogor City Inflation Data, (a) Bogor City Foodstuffs Group Inflation Exploration, (b) ACF and PACF Plot

Based on Figure 3(a), the inflation that occurs in foodstuffs y_1 and y_2 has a relatively small variation compared to inflation in y_3. However, the three foodstuffs did not show an uptrend or a downtrend. Based on the ACF plot in Figure 3(b) shows the ACF, PACF, and CCF plots in descending order. Based on the ACF plot, it can be seen that the y_1 variable has a cuts-off pattern after the k-th lag, while the y_2 and y_3 variables have a tails-off pattern with cosine-shaped. Based on the PACF plot, the three variables show the same tails-off pattern. The CCF plot shows the close relationship between the two variables where the variables y_1 and y_3 do not show any signs of any lag. Meanwhile, between the variables y_2 and y_3, the CCF plot looks significant at the eighth lag, which shows a significant relationship between the two variables after eight periods.

3.2.2 Stationary Test

Figure 3 shows an indication of non stationarity in the PF variable. Furthermore, the stationarity check was carried out with a formal test, the ADF test. The results of the stationary test on the three variables with the ADF test are presented in Table 1.

Table 1. Variable Stationary Test

Variable	**Level**		**Description Stationary**
	***ADF* Test Statistic**	***P- value***	
y_1	-4.1084	0.0213	Not stasionary
y_2	-4.9116	0,01*	Stasionary
y_3	-4.5345	0,01*	stasionary

*Significant at level 0.01

Table 1 shows that the variable y_1 (PF) is not stationary at a significant level of 1%, while y_2 and y_2 are stationary. This means that the inflation data for the Bogor City foodstuff group contains a stationary problem.

3.2.3 Multivariate Time Series Modeling (VAR, VARD, VECM) and Evaluation

The next step is to obtain the best multivariate time series forecasting model in various stationary conditions. Then two modeling scenarios are carried out in various stationary conditions, namely, non-stationary and mixed stationary variables. The next stage is to divide the data into 4-fold. The data for each fold is divided into training data and testing data. Furthermore, three scenarios were modeled and evaluated based on model-fitting and forecasting power. The evaluation results for each fold are presented in Table 2 and Table 3.

Table 2. Evaluation Results of Model Fittings on Four Folds

Fold	RMSE			MAD		
	VAR	VARD	VECM	VAR	VARD	VECM
1st	1.52	1.92	1.56	2.25	3.60	2.38
2nd	1.54	1.94	1.59	2.33	3.70	2.47
3rd	1.60	2.06	1.65	2.53	4.17	2.67
4th	1.73	2.12	1.77	2.95	4.43	3.08

Table 2 presents the results of the evaluation of the model fitting power for three multivariate time series modeling scenarios at 4-fold. The evaluation results indicate that the VAR model is the best. Furthermore, the results of the evaluation of the 4-fold forecasting power are presented in Table 3.

Table 3. Forecasting Power Evaluation Results on Four Folds

Fold	RMSE			MAD		
	VAR	VARD	VECM	VAR	VARD	VECM
1st	1.87	2.55	1.90	3.21	5.55	3.29
2nd	2.48	3.73	2.35	5.48	11.90	5.35
3rd	3.45	2.61	3.31	10.80	6.33	10.60
4th	2.63	4.78	2.58	6.34	19.60	7.28

The results of the evaluation of forecasting power in Table 3 show that there is a difference in the best model, although it appears that the VAR model is still better. Furthermore, the best overall model at 4-fold based on RMSE and MAD was obtained on average. The evaluation results based on fitting power and forecasting power as a whole are presented in Table 4.

Table 4. Overall Fitting Power and Forecasting Power Evaluation Results

Model	Fitting Power		Forecasting Power	
	RMSE	MAD	RMSE	MAD
VAR	1.60	2.51	2.61	6.45
VARD	2.01	3.97	3.42	10.9
VECM	1.64	2.65	2.53	6.64

In Table 4, the power fitting model based on RMSE and MAD is evaluated. Overall the VAR model is obtained as the best model. The evaluation of forecasting power based on MAD obtained the best model, namely the VAR model. Meanwhile, based on RMSE obtained, VECM is the best model. In the case of this study, the VAR model is still chosen as the best model because RMSE is sensitive to outliers. Based on the best 4-fold model, the VAR model is used in the second modeling stage.

VAR modeling at this stage uses all data, starting with determining the optimal lag; the results obtained are lag one. It means that the modeling in the second stage uses the VAR(1) model. Next, estimate the parameters of the VAR(1) model, then form the VAR(1) model. So the VAR(1) model for the inflation data for the Bogor City food group is

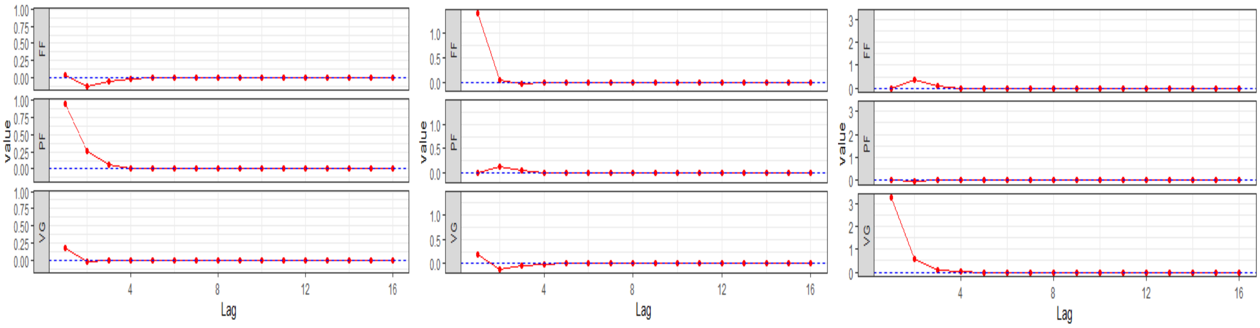
$$\boldsymbol{PF = 0.4078 + 0.2717PF_{t-1} + 0.0914FF_{t-1} - 0.0073VG_{t-1}}$$

$$FF = 0.4865 - 0.1544PF_{t-1} + 0.0155FF_{t-1} + 0.1120VG_{t-1}$$

$$VG = 0.5339 - 0.0526PF_{t-1} + 0.071FF_{t-1} + 0.1080VG_{t-1}$$

The next stage is the diagnostic test on the VAR(1) model; the results obtained indicate that the VAR model has met the assumptions of non-autocorrelation and homoscedasticity. Thus, in this study, it can be concluded that the VAR model is a feasible multivariate time series forecasting model.

The following stage tests the Impulse response function (IRF) and Forecast Error Variance Decomposition (FEVD) on the best multivariate time series forecasting model. IRF was conducted to find out how the shock effect of a variable on the variable itself and other variables. Meanwhile, FEVD estimates how much a variable contributes to the variable itself and other variables in the following several periods, with its value measured in percentage terms. The results of the two tests are presented in Figure 4 and Figure 5.



(a) (b) (c)

Figure 4. IRF on VAR (1) Models, (a) Impulse Response on PF, (b) Impulse Response on FF, (c) Impulse Response on VG

Figure 4(a) shows that the variable PF responded positively to the shock that occurred to oneself, which was quite large at the beginning of the period, and decreased from the second period, then the response given disappeared with increasing time. The FF variable tends to give a negative response to the shock in the PF variable in the initial three periods, and then the given response disappears in the long term. While the VG variable gave a positive response that was not too large at the beginning of the period, the response disappeared with increasing time.

Figure 4(b) shows that the PF variable did not respond tremendously to the shock that occurred in the FF variable in the following two periods. The given response disappeared with increasing time. While the FF variable responded positively to the shock that occurred in the variable itself, which was quite prominent in the initial two periods, the response disappeared with increasing time. Furthermore, the VG variable gave a not too large response to the shock that occurred in the FF variable at the beginning of the period, then gave a slightly negative response in the second period, then the response given disappeared with increasing time.

Figure 4(c) shows that the PF variable did not respond to the shock that occurred in the VG variable from the beginning of the period to the long period. While the FF variable gave a slightly positive response to the shock in the VG variable in the initial two periods, the response given decreased. In the long term, the response disappeared. Next, the best model FEVD test results are presented in Figure 5.

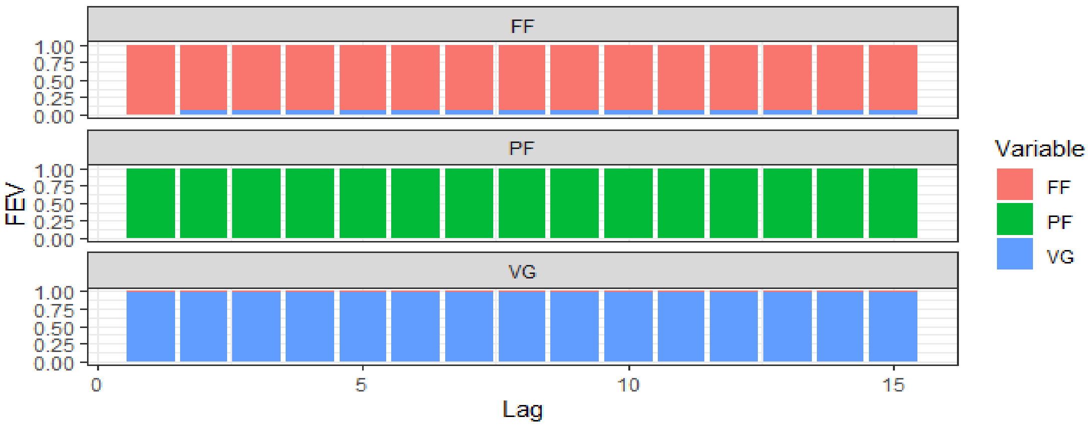
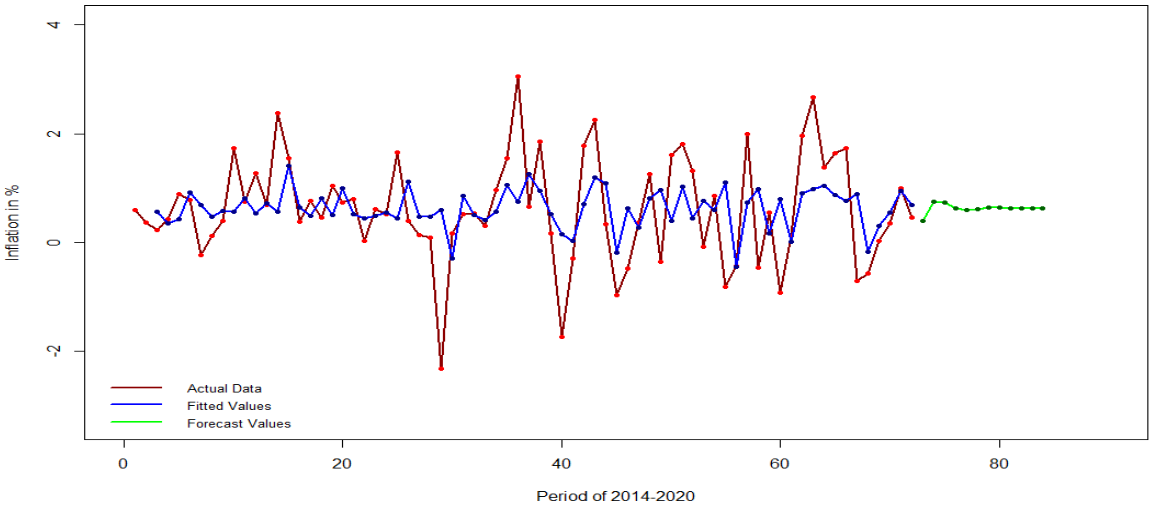
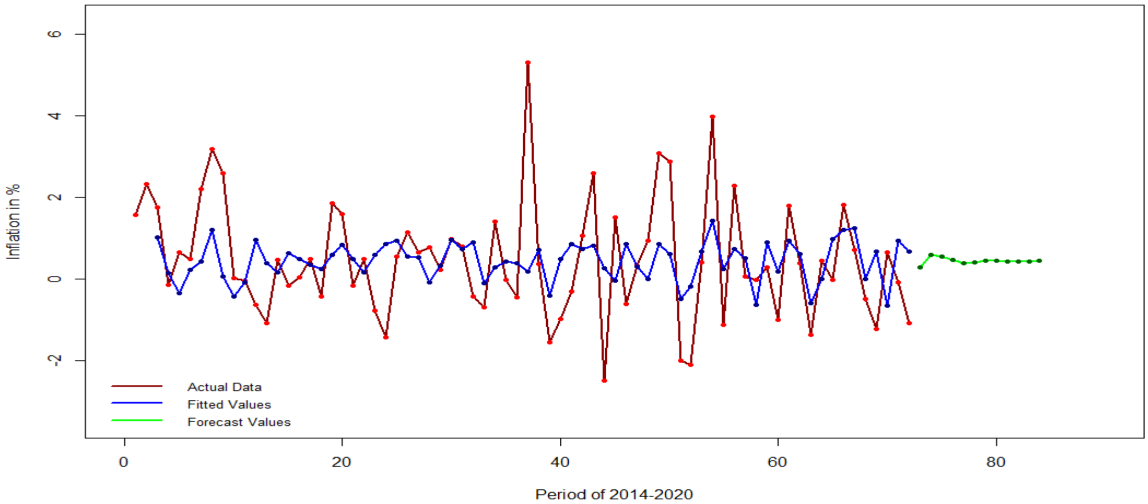


Figure 5. FEVD on VAR (1) Models

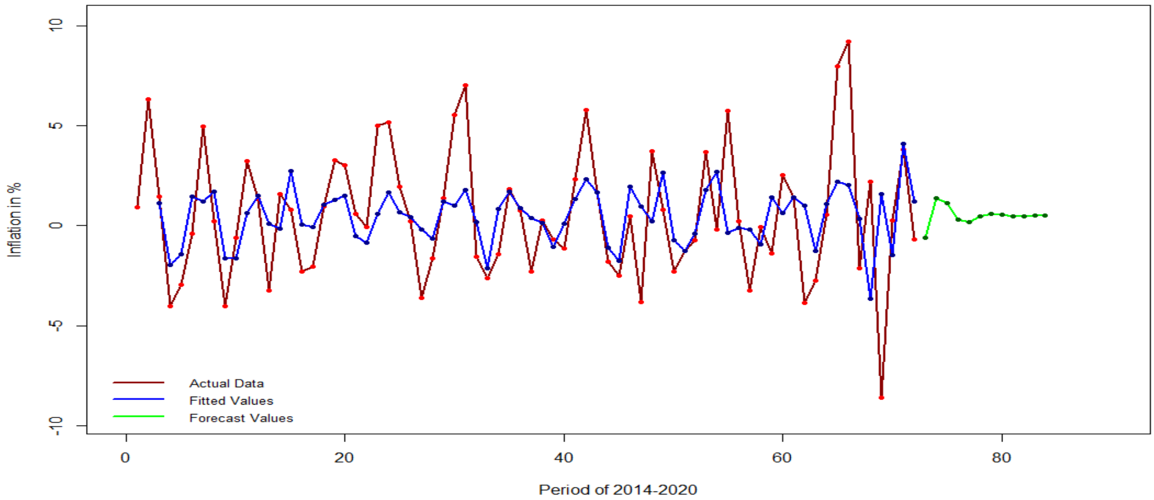
Figure 5 presents the FEVD results for all variables. The first FEVD shows the variables that are estimated to have a significant contribution to the FF variable in the next 15 months, namely the FF variable itself, with an average monthly contribution of more than 95%, followed by the VG variable, which has an average contribution of less than 10%. In contrast, the second FEVD shows the variable estimated to have a significant contribution to the PF variable in the next 15 months, namely the PF variable itself, with an average monthly contribution of almost 100%. Furthermore, the last FEVD shows the variable estimated to significantly contribute to the VG variable in the next 15 months, namely the VG variable itself, followed by the FF variable of less than 5%. The last stage in modeling is forecasting the following few periods using all actual data. The results of the comparison of actual data, estimated values, and forecasting results are presented in Figure 6.



(a)



(b)



(c)

Figure 6. Comparison of Actual Data, Fitted Value and Forecast Value, (a) Comparison of Actual Data, Fitted Value and Forecast Value for PF, (b) Comparison of Actual Data, Fitted Value and Forecast Value for FF, (c) Comparison of Actual Data, Fitted Value and Forecast Value for VG

Based on Figure 6(a), the model estimates the inflation value for the variable PF with a minor variance from the actual data. This can be seen from the width of the plot of the estimated value is smaller than the actual data. However, based on the plot pattern between the estimated value and the actual data, it shows a similarity with the accuracy based on RMSE of 0.9303 and MAD of 0.7071. The forecast results show that inflation in the PF variable will decrease in January 2020 compared to December 2019. Then it will increase in February and March 2020, then experience a slight decrease, and the next inflation forecast will increase steadily in the following few periods.

Figure 6(b) presents a composite plot of FF estimates for inflation with a minor variance from the actual data. This can be seen from the width of the plot of the estimated value is smaller than the actual data. However, based on the plot pattern between the estimated value and the actual data, it shows a similarity with the accuracy based on RMSE of 1.3784 and MAD of 1.0596. The forecast results show that inflation in the FF variable decreased in January 2020 compared to December 2019. Then it increased slightly in February 2020 and then decreased in March 2020, then the next inflation forecast looks constant after several periods ahead.

Figure 6(c) presents a combined VG plot of estimated inflation values with a minor variance compared to the actual data. This can be seen from the width of the plot of the estimated value is smaller than the actual data. Based on the plot form, the estimated value and the actual data still show similarities, although not precisely the same, with an accuracy based on RMSE of 3.1822 and MAD of 2.4810. The results of the inflation forecast showed a decline in January 2020 compared to December 2019, then experienced a relatively high increase in February. The next inflation forecast will decline in the next three months, then be seen to be constant in the following several periods.

4. CONCLUSIONS

In this study, a superior multivariate time series forecasting model was obtained based on the model's goodness in various stationary conditions. The evaluation of the multivariate time series model on simulation modeling in various stationary conditions is that the vector autoregressive (VAR) model is the best model. This is indicated by the stability of the performance of the VAR model in 100 replications. Evaluating inflation data modeling in Bogor City with mixed stationary conditions showed that the VAR model was the best multivariate time series, forecasting model. This is based on the value of two measures of model goodness for evaluating model fitting power and forecasting power. The multivariate time series model that is formed on the inflation of the foodstuff group in Bogor City is the VAR(1) model. This means the VAR model is good enough to be used as a forecasting model in mixed stationary conditions. Thus, in this study, based on the goodness of the model in two modeling scenarios in various stationary conditions, overall, it was found that the VAR model was superior to the VARD and VECM models. The limitation of this study is that the outlier effect is not handled further, so for future studies, it is possible to add outlier effects and increase the number of replications in the simulation modeling.

REFERENCES

[1] W. Enders, *Applied Econometric Time series*, 4th ed. USA: J Wiley, 2014.

[2] A. Salim and Fadilla, "Pengaruh Inflasi Terhadap Pertumbuhan Ekonomi Indonesia Anggun Purnamasari," *Ekon. Sharia J. Pemikir. dan Pengemb. Ekon. Syariah*, vol. 7, no. 1, pp. 17–28, 2021, [Online]. Available: www.bps.go.id,.

[3] E. Pranyoto, "Analisis Kointegrasi dan Kausalitas Engel Granger Tingkat Suku Bunga Simpanan , Perubahan Nilai Tukar Rp/USD dan return Pasar Saham Di Bursa Efek," *J. Bisinis Darmajaya*, vol. 03, no. 02, 2017.

[4] Iskandar, "Analisis Vector Autoregression (VAR) Terhadap Interrelationship Antara Financing Deposit Ratio (FDR) DAN Return ON Asset (ROA) Pada Bank Syariah Di Indonesia," *J. JESKape*, vol. 3, no. 2, pp. 19–39, 2019.

[5] Erwandi, F. M. Afendi, and B. Waryanto, "Analisis Pengaruh Daerah Pemasok Terhadap Harga Cabai Merah Di Dki Jakarta Menggunakan Vector Error Correction Model (Vecm)," *Indones. J. Stat. Its Appl.*, vol. 3, no. 3, pp. 216–235, 2019, doi: 10.29244/ijsa.v3i3.276.

[6] L. M. Hamzah, S. U. Nabilah, E. Russel, M. Usman, E. Virginia, and Wamiliana, "Dynamic Modeling and Forecasting of Data Export of Agricultural Commodity by Vector Autoregressive Model," *J. Southwest Jiatong Univ.*, vol. 55, no. 3, 2020.

[7] S. Aktivani, "Uji Stationeritas Data Inflasi Kota Padang Periode 2014-2019," *J. Stat. Ind. dan Komputasi*, vol. 6, no. 1, pp. 26–33, 2021.

[8] M. R. Susila, "Pemodelan Multivariat Kunjungan Wisatawan Dan Darat Yang Melibatkan Dampak Wabah Covid-19," vol. 15, no. 3, pp. 467–478, 2021.

[9] R. H. Assaad and S. Fayek, "Predicting the Price of Crude Oil and its Fluctuations Using Computational Econometrics: Deep Learning, LSTM, and Convolutional Neural Networks," *Econom. Res. Financ.*, vol. 6, no. 2, pp. 119–137, 2021, doi: 10.2478/erfin-2021-0006.

[10] P. A. V. B. Swamy and P. von zur Muehlen, "Cointegration: Its fatal flaw and a proposed solution," *Sustain. Futur.*, vol. 2, p. 100038, 2020, doi: 10.1016/j.sftr.2020.100038.

[11] J. Van Greunen, A. Heymans, C. Van Heerden, and G. Van Vuuren, "The prominence of stationarity in time series forecasting," *J. Stud. Econ. Econom.*, vol. 38, no. 1, pp. 1–16, 2014, doi: 10.1080/10800379.2014.12097260.

[12] F. Bashir and H. L. Wei, "Handling missing data in multivariate time series using a vector autoregressive model-imputation (VAR-IM) algorithm," *Neurocomputing*, vol. 000, pp. 1–8, 2017, doi: 10.1016/j.neucom.2017.03.097.

[13] E. Zivot and J. Wang, "Modeling Financial Time Series with S-PLUS," *J. R. Stat. Soc. Ser. A (Statistics Soc).*, 2016, doi: 10.1111/j.1467-985x.2007.00506_20.x.

[14] K. Bulteel, F. Tuerlinckx, A. Brose, and E. Ceulemans, "Clustering vector autoregressive models: Capturing qualitative differences in within-person dynamics," *Front. Psychol.*, vol. 7, no. OCT, pp. 1–15, 2016, doi: 10.3389/fpsyg.2016.01540.

[15] A. Suharsono, A. Aziza, and W. Pramesti, "Comparison of vector autoregressive (VAR) and vector error correction models (VECM) for index of ASEAN stock price," in *AIP Conference Proceedings*, 2017, vol. 1913, doi: 10.1063/1.5016666.

[16] R. M. Ueda, A. M. Souza, and R. M. C. P. Menezes, "How macroeconomic variables affect admission and dismissal in the Brazilian electro-electronic sector: A VAR-based model and cluster analysis," *Phys. A Stat. Mech. its Appl.*, vol. 557, p. 124872, 2020, doi: 10.1016/j.physa.2020.124872.

[17] E. G. Elizabeth, C. S. Nicole, K. S. Damian, and F. S. Anne, "Using Vector Autoregression Modeling to Reveal Bidirectional Relationships in Gender/Sex-Related Interactions in Mother–Infant Dyads," *Front. Psychol.*, vol. 11, no. August, pp. 1–19, 2020, doi: 10.3389/fpsyg.2020.01507.

[18] T. Chai and R. R. Draxler, "Root mean square error (RMSE) or mean absolute error (MAE)? -Arguments against avoiding RMSE in the literature," *Geosci. Model Dev.*, vol. 7, no. 3, pp. 1247–1250, 2014, doi: 10.5194/gmd-7-1247-2014.

[19] U. Khair, H. Fahmi, S. Al Hakim, and R. Rahim, "Forecasting Error Calculation with Mean Absolute Deviation and Mean Absolute Percentage Error," *J. Phys. Conf. Ser.*, vol. 930, no. 1, 2017, doi: 10.1088/1742-6596/930/1/012002.

[20] J. Yuan, Y. Xu, and X. Zhang, "Income Growth, Energy Consumption, and Carbon Emissions: The Case of China," *Emerg. Mark. Financ. Trade*, vol. 50, no. 5, pp. 169–181, 2014, doi: 10.2753/REE1540-496X500512.