Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 124 (2017) 189–196

Procedia Computer Science

www.elsevier.com/locate/procedia

4th Information Systems International Conference 2017, ISICO 2017, 6-8 November 2017, Bali, Indonesia

The Performance of ARIMAX Model and Vector Autoregressive (VAR) Model in Forecasting Strategic Commodity Price in Indonesia

Wiwik Anggraeni[a]*, Kuntoro Boga Andri[b], Sumaryanto[c], Faizal Mahananto[a]

[a] *Sepuluh Nopember Institute of Tehnology, Faculty of Information Technology, Department of Information Systems, Jl. Arief Rahman Hakim, Surabaya 60111, Indonesia*

[b] *Agriculture Technology Assessment Center, Riau 65163, Indonesia*

[c] *Center of Economic Study and Agriculture Policy, Bogor, Indonesia*

Abstract

Rice as one of the strategic commodities has an important role in the life of Indonesian society. This is cause of rice is the main food of the Indonesian nation. Therefore, the stabilization of food prices is one of the priorities of the Indonesian government's policy. It can minimize the impact of the global financial crisis such as inflation and purchasing power of the poor. The sta bility price can be maintained by price forecasting for several periods ahead. It can be used to set up the anticipatory action. In this research, ARIMAX model and VAR model used to forecast the rice price. This model involves several variables including consumer rice price (HKB), production (PROD), dry milled rice (GKP), harvested area (LP), and rice price in Thailand (HD). The results show that ARIMAX model can predict the rice consumer price with MAPE 0.15%. This is 15.27 % better than VAR model. The GKP variable did not significantly affect to the rice price. This is indicated by the MAPE difference between model with GKP and model without GKP is less than 0.01%.

Peer-review under responsibility of the scientific committee of the 4th Information Systems International Conference 2017.

Keywords: Forecasting; Arimax; Vector Autoregressive; Price; Strategic Commodity; Rice

* Corresponding author. Tel.: +6-231-599-9944; fax: +6-231-596-4965.
E-mail address: wiwik@is.its.ac.id

1877-0509
Peer-review under responsibility of the scientific committee of the 4th Information Systems International Conference 2017
10.1016/j.procs.2017.12.146

1. Introduction

Rice has an important role in the life of Indonesian society. It is considered an important food because it is the main staple food of the Indonesia. Rice consumption in Indonesia is among the highest in the world [1]. The average national rice consumption in 2010 was 139.15 kg per capita / year and corrected to 113.48 kg per capita / year 2011 compared to the consumption rate of rice in some neighboring countries (Malaysia: 80 kg / person / year, Thailand: 90 kg / person / year and Japan: 60 kg / person / year) or the average world per capita consumption of rice only 60 kg per capita / year. Therefore, the supply of rice becomes very important. Rice supply in Indonesia also has an influence on several areas, such as economic, environmental, and socio-political fields [1]. Most Indonesians want a stable supply and price of rice, available over time, distributed evenly, and provide affordable prices [2].

Seeing the importance of the need and influence of rice, the Government of Indonesia always strives to improve food security, one of which is to maintain rice supplies [1]. Sufficient rice content will also affect the price of rice on the market. The price of rice also has a very significant influence on the various areas as to which the supply of rice [3]. In addition to rice supplies, the price on the market is influenced by other things such as rice production, harvested area, rice prices in other countries, and others. For that reason, stabilizing rice prices has become one of the priorities of Indonesian government policy, as this can minimize the impact of the Indonesian financial crisis as inflation and purchasing power of the poor [4]. The stability price can be maintained by price forecasting for several periods ahead. It can be used to set up the anticipatory action [5].

The VAR forecasting model is widely used by researchers and shows success in forecasting macroeconomic and regional variables [6]. VAR was once used to build models by looking at relationships between variables by determining co integration vectors in the variables used [7] and for predicting food consumption, the Consumer Price Index (CPI), Gross Domestic Product (GDP), and investment in South Africa [8] . In addition, VAR is also used to predict macroeconomic data [9], government bond yields [10] and point and density in Euro [11].

While the ARIMAX model has not been widely used in terms of forecasting in the economic field, it proves good to predict variables that are affected by some other variables. ARIMAX proves to be good for modeling the growth of influenza disease [12], kids clothes demand forecasting [13], forecasting number of dengue fever cases [14], and forecasting number of tuberculosis patients [15].

Based on this,in this research, ARIMAX model is tried to be applied in forecasting economic data (price of rice) which have fluctuation and level of uncertainty is very high. After that, the results are compared with forecasting results using the VAR model.

2. Data description

The data which is used is in the form of monthly period data from 2000-2015. The variables used are retail consumer price (in units of Rupiah/kg), rice production (in tons), domestic and overseas rice procurement (in ton, net), dry paddy harvest price (Rupiah/kg), rice harvest area (in hectares), Bangkok rice price 5% (in US $/MT, FOB). The interpolation method is used to complete the blank entries in the retail consumer price rice data. Meanwhile, triple exponential smoothing is used to complete the data of rice procurement, the price of dry grain harvest, and harvested area. Each variable has a different data pattern. The pattern of each data is shown by Fig. 1.

In this study, retail consumer price rice acts as an independent variable, while rice production, rice procurement, dry harvested paddy price, harvested area, world rice price (Bangkok 5%) act as dependent variable.

3. Methodology

The methodology adopted for forecasting is as follows.

3.1. Stationary data test

The result of the stationer test in the variant shows that the price of rice consumer (HKB) is not stationary in Box Cox variance. Plot is shown in picture 2 that the rounded value obtained from the box-plot check is 0.001 with the

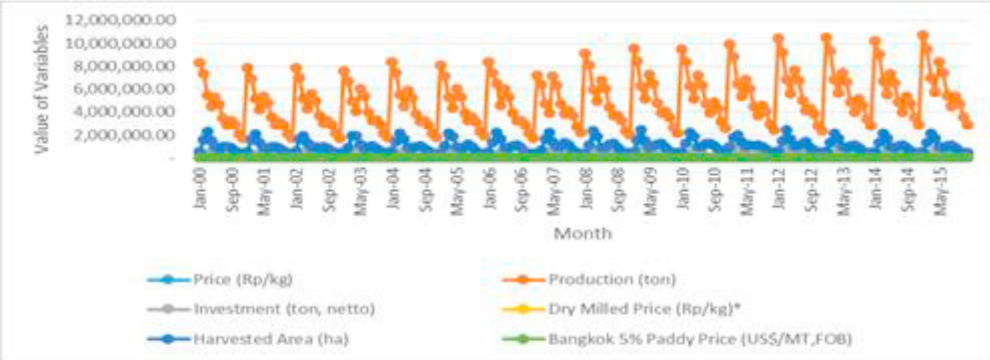


Fig. 1. Plots of Variables

lower limit and the successive upper limit of -0.26 and 0.38, based on the rounded value of 0.25 and interval intervals that do not pass the value.

Since the data is not stationary, the next data must be transformed in order to make the data stationary in variance. Of the several types of transformations performed on this data, it turns out the log transformation with the zero lambda value the most suitable. After stationary in the next variance, the transformed data also needs to be tested stationary in averages using Augmented Dickey-Fuller test [16][17]. The result of unit root test indicates that the data is not stationary in the mean. The unit root test results are shown by Fig. 2 and Fig. 3.

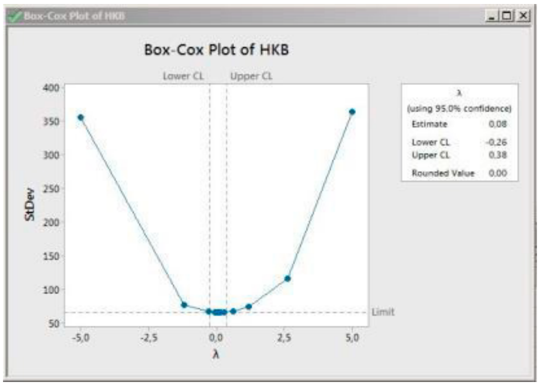


Fig. 2. Box-Cox Plot Result

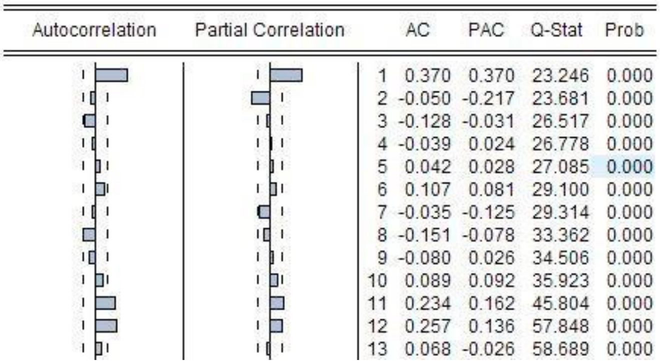
Fig. 3. Augmented Dickey-Fuller Result

3.2. Differencing

Differencing is required when the data is not stationary in the mean. The number of differencing processes depends on how many processes are performed to make the data stationary. The process o f differencing is done by the difference between data at t period and t+1 period [16]. After the data is differentiated, it turns out the data is stationary. This is evident from the augmented dickey-fuller test which shows that probability of <0.05 and | t-statistic |> | t-critical value |.

3.3. Estimated ARIMA and VAR parameters

Parameter estimation is done by looking at plot Auto-correlation Function (ACF) and Partial Auto-correlation Function (PACF) data [16][17]. Estimates are conducted to determine whether the model used AR, ARMA, or ARIMA and specify the order of each model. Pieces of ACF and PACF results are demonstrated by Fig. 4.



Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.370	0.370	23.246	0.000
		2	-0.050	-0.217	23.681	0.000
		3	-0.128	-0.031	26.517	0.000
		4	-0.039	0.024	26.778	0.000
		5	0.042	0.028	27.085	0.000
		6	0.107	0.081	29.100	0.000
		7	-0.035	-0.125	29.314	0.000
		8	-0.151	-0.078	33.362	0.000
		9	-0.080	0.026	34.506	0.000
		10	0.089	0.092	35.923	0.000
		11	0.234	0.162	45.804	0.000
		12	0.257	0.136	57.848	0.000
		13	0.068	-0.026	58.689	0.000

Fig. 4. ACF and PACF Plot

The result of the model guess is ARIMA (0, 1, 1), ARIMA (0, 1, 0), ARIMA (1, 0, 0), ARIMA (1, 1, 1), ARIMA (2, 1, 2), ARIMA (2, 1, 2), ARIMA (2, 1, 1). As for VAR, to find out the minimum number of co - integration vectors, the Johansen co-integration test (maximum Eigen value) is used. The test results are illustrated in Table 1. the results depicts that there are at least 6 co-integration vectors that can be formed.

Table 1. Johansen Co-Integration Test Result

Hipotesis	Eigenvalue	Max-Eigen Statistic	Critical Value α=5%	Prob.**
None*	0,709230	187,7539	40,07757	0,0001
1*	0,580009	131,8635	33,87687	0,0000
2*	0,424718	84,04005	27,58434	0,0000
3*	0,304216	55,13292	21,13162	0,0000
4*	0,258705	45,50229	14,26460	0,0000
5*	0,226189	38,97705	3,841466	0,0000

3.4. ARIMA significance test

All ARIMA models obtained from the estimation process which will be tested for its significance by looking at the probability values of the model. If the probability value of all variables ⩽ 0.05 and | t-statistic | all variables> t-table, then the model is said to be significant and diagnostic test can be done [16][17]. The significance test results for the guess models are demonstrated in Table 2.

Table 2. Significance test result

Model	Status	Model	Status	Model	Status
ARIMA (0, 1, 1)	Significant	ARIMA (2, 1, 1)	Insignificant	ARIMA (1, 2, 1)	Insignificant
ARIMA (0, 1, 2)	Insignificant	ARIMA (2, 1, 2)	Significant	ARIMA (1, 2, 2)	Significant
ARIMA (1, 1, 0)	Significant	ARIMA (0, 2, 1)	Significant	ARIMA (2, 2, 0)	Significant
ARIMA (1, 1, 1)	Insignificant	ARIMA (0, 2, 2)	Significant	ARIMA (2, 2, 1)	Significant
ARIMA (1, 1, 2)	Significant	ARIMA (1, 2, 0)	Significant	ARIMA (2, 2, 2)	Insignificant
ARIMA (2, 1, 0)	Insignificant				

3.5. Diagnostic test

From 10 significant models performed ARIMA diagnostic test to test the feasibility of the model when viewed from the remaining sneeze and the homogeneity of the residual, the ARIMA model is said to be feasible if the correlogram-Q statistics function shows a p> 0.05 indicating that the remainder does not have a specific pattern (random) and the squared residuals' correlogram notes p> 0.05 which denotes the model's remaining homogeneity[16][17].

3.6. Randomness and homogeneity test

Table 3 shows the results of randomness and homogeneity test.

Table 3. Randomness and homogeneity test result

Model	Randomness Test	Homogeneity Test	Result
ARIMA (0,1,1)	Passed	Passed	Passed
ARIMA (1,1,0)	Didn't pass	Passed	Didn't pass
ARIMA (1,1,2)	Passed	Passed	Passed
ARIMA (2,1,2)	Didn't pass	Passed	Didn't pass
ARIMA (0,2,1)	Didn't pass	Passed	Didn't pass
ARIMA (0,2,2)	Didn't pass	Didn't pass	Didn't pass
ARIMA (1,2,0)	Didn't pass	Didn't pass	Didn't pass
ARIMA (1,2,2)	Didn't pass	Passed	Didn't pass
ARIMA (2,2,0)	Didn't pass	Didn't pass	Didn't pass
ARIMA (2,2,1)	Didn't pass	Passed	Didn't pass

3.7. Selection of the best ARIMA model

Once a model has passed from all tests, the next step is to select the best model among all test -escaped models. The selection of the best model of ARIMA is done by using the Akaike Information Criteria (AIC) and Schwarz Information Criteria (SIC) values obtained such as when performing the significance test [16][17]. The AIC and SIC values of the escaped model are illustrated in Table 4. From Table 4, it is seen that ARIMA (0, 1, 1) has the smallest AIC and SIC values, so the ARIMA (0, 1, 1) model selected for ARIMAX process.

Table 4. AIC and SIC Value of Model

	ARIMA (0, 1, 1)	ARIMA (1, 1, 2)
AIC	-5.002154	-5.001298
SIC	-4.964812	-4.945286

Tabel 3 shows that ARIMA(0, 1, 1) has the smallest AIC and SIC value. Therefore ARIMA(0, 1, 1) model was choosing for further process.

3.8. Heteroskedasticity test

After obtaining the best model of ARIMA,the Heteroskedasticity test is used to test whether the variant is influenced by the preceding squares previously and the previous variant [16][17].

3.9. Granger causality test

This test is used in the VAR model. The causality test indicates whether a variable has a two -way or one-way relationship. In ARIMAX, causality between variables are checked by linearity and multicollinearity Test [16][17]. Granger causality test is used to determine the relationship between one variable with other variables that affect each other. The entire causal relationship between the above variables can be illustrated in Table 5. The causality test does not show a direct cause-and-effect relationship between rice production variables and rice consumer prices, but forecasting with these two variables needs to be considered because it proves to have the smallest error rate.

Table 5. Relationship between variables

Dependent	GKP	HD	HJ	HKB	LP	PI	PROD
GKP		No	No	No	No	No	Yes
HD	No		No	No	No	No	No
HJ	No	No		No	No	No	No
HKB	Yes	No	No		No	No	No
LP	No	No	No	No		No	Yes
PI	Yes	No	No	No	Yes		No
PROD	No	No	No	No	Yes	Yes	

3.10. Modelling and forecasting

After getting the best model ARIMA and it passes Heteroskedastisitas test, then the next is doing ARIMAX modelling. In ARIMAX modeling, it was done by using the best ARIMA model and then process it for forecasting. The preferred method is LS - Least Square (NLS and ARMA) method. Here, it creates several different models based on the available variables and correlation test between the free variables with the dependent variable.

Likewise for VAR, forecasting model is formed based on the image x. in this study, they are formed 25 models but after getting tested, it was not all produce good forecasting.

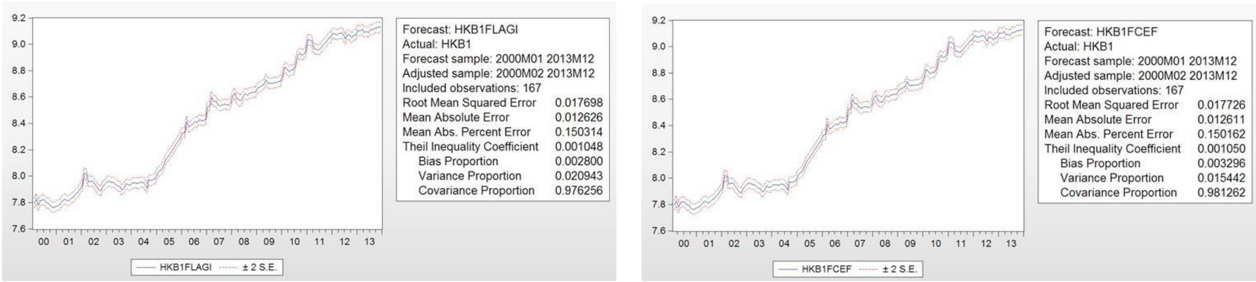
4. Results and discussions

ARIMAX model built here is 2. Model (1) is a model to predict HKB based on PROD, GKP, LP, and HD variables. Model (2) is a model to forecast HKB based on PROD, LP, and HD variables. The result of forecasting for model (1) and model (2) is demonstrated in Fig. 5(a) and Fig. 5(b). The performance of ARIMAX model is shown by Table 6. As for VAR, the best five models from each group are model 5 (models involving HKB and PROD, model 14 (models involving HKB and PROD, model 23 (models involving HKB, GKP, HD, LP, and PROD), model 24 (models involving HKB, HD, LP, PI, and PROD), model 25 (models involving HKB, GKP, PI, LP, HD, and PROD). For the forecasting results of each model is shown by Fig. 6. On the other hand, for performance results for models VAR is shown in Table 7.

5. Conclusions

The VAR model forecasting rice prices has not been able to provide maximum results. VECM methods are required to provide better results. Consumer price determination is also influenced by other factors or variables. Based on the test of error rate on each model, the most influencing variable data on consumer price of rice is the variable of previous rice price and rice production.

Although on the causality test of historical data, rice production variables are not related to the price of rice, but according to experts and informants the production variables are expressed to have a relationship to the price of rice. The determination of rice price that does not involve rice production variables has a greater error rate than that involving rice production variables.



(a) (b)

Fig. 5. Plots of Forecasting Result using (a) modul 1 and (b) modul 2.

Table 6. Performance of ARIMAX Model

ERROR	DATA TRAINING		DATA TESTING	
	MODEL 1	MODEL 2	MODEL 1	MODEL 2
MAPE	1.26%	1.26%	1.91%	1.92%
MSE	0.03%	0.03%	0.05%	0.05%

Fig. 6. Plots of Forecasting Result

Table 7. Performance of VAR Model

Model	Variabel	MAPE (Training	MAPE (Testing
5	HKB, PROD	23,28%	2,30%
14	HKB, LP, PROD	23,14%	2,30%
20	HKB, HD, LP, PROD	23,15%	2,30%
23	HKB, GKP, HD, LP, PROD	22,21%	2,33%
24	HKB, HD, LP, PI, PROD	21,77%	2,41%
25	HKB, GKP, PROD, PI, LP, HD	21,39%	2,44%

The results show that ARIMAX model can predict the paddy consumer price with MAPE 0.15%. This is 15.27% better than VAR model. The GKP variable did not significantly affect the paddy price. This is indicated by the MAPE difference between model with GKP and model without GKP is less than 0.01%. In the same time, it can be said in general that ARIMAX model can predict the paddy consumer price with MAPE 0.15%. This is 15.27% better than VAR model.

In terms of MAPE, this VAR model can still be said to be good (less than 10%), but forecasting results that are shown graphically depicts the difference between historical data and data forecasting results. For the next one, it needs to be analyzed why it can happen and whether there are factors that cause it.

Acknowledgements

The research was funded by The Research and Development Agency on Agriculture, Indonesian Agriculture Ministry under the 2017 Strategic Research, Assessment and Strategic Development (KP4S) scheme.

References

[1] BULOG, "Ketahanan Pangan," BULOG, 2014. [Online]. Available: http://www.bulog.co.id/ketahananpangan.php. [Accessed 8 December 2015].
[2] B. -. K. Pertanian, "Ekonomi Beras Indonesia, Peranan Beras dalam Perekonomian Nasional," BB Padi (Balai Besar Penelitian Tanaman Padi), 2 Oktober 2015. [Online]. Available: http://bbpadi.litbang.pertanian.go.id/index.php/berita/berita-utama/content/216- ekonomi-beras-indonesia-peranan-beras-dalam-perekonomian-nasional. [Accessed 8 December 2015].
[3] David, D. Timmer, P. Why Stable Food Prices are a Good Thing: Lessons From Stabilizing Rice Price in Asia . Global Food Security Journal. 2012. 127-133
[4] C. T. David Dawe. Why stable food prices are a good thing : Lessons from stabilizing, Global Food Security. 127-133, 2012.
[5] F. A. MARJUKI. Analisis faktor faktor yang mempengaruhi harga beras di indonesia tahun 1981-2006. 2008.
[6] D. E. Spencer, Developing a Bayesian vector autoregression forecasting model, International Journal of Forecasting, 9: 407-421, 1993.
[7] J. F. D. G. D. Lixin SUN, Bank loans and the effects of monetary policy in China: VAR/VECM approach. China Economic Review. 21; 65-97. 2010.
[8] R. Gupta, FORECASTING THE SOUTH AFRICAN ECONOMY WITH," openUP, 2007.
[9] R. G. K. K. Mehmet Balcilar. Forecasting macroeconomic data for an emerging market with a nonlinear. Economic Modelling. 44: 215-228. 2015.
[10] G. K. M. M. Andrea Carriero. Forecasting government bond yields with large Bayesian Vector Autoregressions. Journal of Banking & Finance.36: 2026-2047. 2012.
[11] S. R. H. Tim O. Berg. "Point and density forecasts for the euro area using. International Journal of Forecasting. 2015; 31: 1067-1095
[12] S. Chadsuthi, S. Iamsirithaworn, W. Triampo and C. Modchang. Modeling seasonal influenza transmission and its association with climate factors in Thailand using Time-Series and ARIMAX analysis. 2015
[13] Anggraeni, W., Vinarti, R.A., and Kurniawati,Y.D. Performance comparisons between ARIMA and ARIMAX method in moslem kids clothes demand forecasting: case study. Procedia Computer Science. 2015; 72: 630-637
[14] Anggraeni, W., Aristiani, L.Using google trend data in forecasting number of dengue fever cases with ARIMAX method case study: Surabaya, Indonesia. 2016 International Conference on Information, Communication Technology and System, Surabaya. 2016; 114-118.
[15] Anggraeni, W., Pusparindra, N., Riksakomara, E., Samopa, F., Pujiadi. The Performance of ARIMAX method in forecasting number of tuberculosis patients in Malang Regency, Indonesia, Indonesia. International Journal of Applied Engineering Research. 2017; 17: 6806-6813.
[16] Spyros Makridakis, Steven C. Wheelwright, and Rob J. Hyndman, Forecasting: Methods and Applications, 3rd ed.: John Wiley & Sons, 1998
[17] O'connell Bowerman and Koehler , Forecasting, Time Series, and Regression, 4th ed. United States of America: Curt Hinrichs, 2005