BAREKENG: Journal of Mathematics and Its Applications
March 2023 Volume 17 Issue 1 Page 0253–0262
P-ISSN: 1978-7227 E-ISSN: 2615-3017
https://doi.org/10.30598/barekengvol17iss1pp0253-0262

SIMULATION OF THE SARIMA MODEL WITH THREE-WAY ANOVA AND ITS APPLICATION IN FORECASTING LARGE CHILLIES PRICES IN FIVE PROVINCES ON JAVA ISLAND

Ratna Nur Mustika Sanusi [1], Budi Susetyo [2*], Utami Dyah Syafitri [3]

[1,2,3] *Department of Statistics, Faculty of Mathematics and Natural Science, IPB University*
Jl.. Raya Dramaga Kampus IPB Dramaga Bogor 16680 West Java, Indonesia

Corresponding author e-mail: * **budisu@apps.ipb.ac.id**

Article History:

Received: 30th September 2022
Revised: 18th December 2022
Accepted: 29th January 2023

Keywords:

Experimental design;
Outliers;
Seasonal Arima;
Simulation;
Three-way ANOVA.

ABSTRACT

Commodities that become potential in the Horticulture Sub-sector are large chilies, so supply and prices must be controlled. One of the efforts that can be made is to predict the price of large chili in the future. However, forecasting is sometimes constrained by several things, such as small sample sizes and outliers. The effect of several factors on the parameter estimation bias can be determined by experimental design by simulating the data obtained from the generation results with several scenarios. The results of the analysis show that all factors have a significant effect on the magnitude of the parameter bias, so all factors can affect forecasting results. When applying forecasting methods to actual data, paying attention to these three factors is necessary. The application of actual data using the SARIMA method gives good results. It can be seen from the RMSE and MAPE values , which tend to be small. Based on the forecast results for the following 12 periods, it is estimated that the price of big chili in 2022 in five provinces will still fluctuate. The high price of chili in five provinces is predicted to reach its highest in the first three months of 2022. The highest price is predicted to occur in DIY Province in February, which is Rp. 74.230.00/kg. However, from the middle to the end of the year, prices will tend to fall and stabilize. The price will be the lowest in Middle Java Province in December, which is Rp. 20,689.00/Kg.

This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution-ShareAlike 4.0 International License.

How to cite this article:

R. N. M. Sanusi, B. Susetyo and U. D. Syafitri., “SIMULATION OF THE SARIMA MODEL WITH THREE-WAY ANOVA AND ITS APPLICATION IN FORECASTING LARGE CHILLIES PRICES IN FIVE PROVINCES ON JAVA ISLAND,” *BAREKENG: J. Math. & App.*, vol. 17, iss. 1, pp. 0253-0262, March 2023.

Copyright © 2023 Author(s)
Journal homepage: *https://ojs3.unpatti.ac.id/index.php/barekeng/*
Journal e-mail: *barekeng.math@yahoo.com; barekeng.journal@mail.unpatti.ac.id*

Research Article • Open Access

1. INTRODUCTION

Commodities that have potential in the Horticulture Sub-sector are large chilies which are included as staples, so their availability and price must be controlled [1]. However, so far, the price of chili commodities is still very fluctuating from year to year in each region, so a data analysis tool is needed to help overcome this problem. One of the efforts from a statistical point of view is to forecast the price of large chilies in the future [2].

Forecasting chili prices using the Seasonal Autoregressive Integrated Moving Average (SARIMA) method has been carried out in Kolaka with the result that the SARIMA method is a suitable method for predicting red chili prices [3]. Besides being used to predict chili prices, the SARIMA method can also be used in other fields such as health, traffic, tourism, economics, etc. Research in the health sector using SARIMA has been carried out to examine the development of Covid-19 cases in Thailand [4]. Research in the field of traffic has also been carried out to examine the incidence of traffic accidents in Belgrade [5]. This research shows that the SARIMA method is still a popular method to be used as a forecasting method. Although it is a classic method, the SARIMA method in tourism provides a higher level of forecasting accuracy than the Support Vector Regression and Fuzzy Time Series methods [6],[7].

Based on the benefits provided by the SARIMA method in several studies that have been carried out, it is necessary to have further studies related to the SARIMA method. This study aims to determine the characteristics of the SARIMA method and if there are any that affect the results of the analysis. These factors, such as the small number of available samples [8] and the pencil value, can cause bias in the estimation parameters [9]. The influence of several factors on the bias of the estimation parameters can be known with certainty by conducting an experimental design. The experimental design generally provides an efficient way to estimate the effect of changing the input model on its output [10]. The experimental design can be done by simulating data generation with a predetermined generation.

In this study, the experimental value design was carried out with three factors: the parameter variation factor, the variation factor in the number of pencil values, and the sample size variation factor. The results of the experimental design of three factors are used as the basic theory for forecasting the price of large chilies in five provinces on the island of Java.

2. RESEARCH METHODS

This study uses two data: generation data for simulation and actual data. The simulation process is carried out to determine the consistency of model accuracy and the interaction of the three factors. The level of consistency of model accuracy can be known by looking at the average RMSE [11]. In contrast, the interaction of 3 factors can be known by looking at the analysis table of variance [12]. The simulation process is carried out based on the ARIMA$(1,0,0)(1,0,0)^{12}$ model using 84, 204, and 504 time series data lengths (n). The data generation process was carried out five times in each scenario. These five generations represent five location variables. There are four scenarios in the simulation process: A, B, C, and D. Each scenario contains an outlier value of 0%, 2.5%, 5%, and 7.5% for each generation. The data generation and modeling process was repeated 100 times, while the experiment design was repeated 20 times for each treatment.

The actual data used in this study is monthly data that includes price variables in the provinces of East Java, Central Java, DI Yogyakarta, West Java, and Banten. The data source in this study was obtained from the Basic Needs Market Monitoring System (SP2KP) of the Ministry of Trade from January 2015 to December 2021. The data is divided into two parts: training data and testing data. Training data is used to create a tentative model consisting of January 2015 to December 2020, while testing data is used to validate the best model from January 2021 to December 2021.

2.1 Stages of Simulation and Applying SARIMA

The stages of data analysis carried out in this study are as follows:

1. Determine the basic model used as the basis for data generation.
2. Determine the length of time series data (n) as many as 84, 204, and 504 data.

3. Generating seasonal time series data based on a predetermined seasonal time series model, namely ARIMA(1,0,0)(1,0,0)^{12}, with 100 replications.
4. Determine the size of the non-seasonal and seasonal parameters for the five seasonal time series models by randomly generating 10 data, with min = -1 and max = 1. The following is the result of generating parameter values for the five locations:
   a. Model 1 : ϕ=0,62 and Φ=0,30
   b. Model 2 : ϕ=-0,17 and Φ=0,24
   c. Model 3 : ϕ=0,42 and Φ=-0,14
   d. Model 4 : ϕ=-0,71 and Φ=0,52
   e. Model 5 : ϕ=-0,69 and Φ=-0,19
5. Add a constant value of 1000 so that the generated data is positive.
6. Perform four generation scenarios, namely scenarios A, B, C, and D, with a different number of outliers inserted in each scenario. The four scenarios each have a total outlier value of 0%, 2.5%, 5%, and 7.5%.
7. Determine the outlier value with the formula ω = 1.5k with k= max – min in each series.
8. Inserting outliers in each series. The outlier point is selected by dividing $\frac{1}{n+1}$ and making it an interval between outliers [13].
9. We are modeling all scenarios of the simulation data using the SARIMA method.
10. Calculate the average of 100 estimators of parameters ϕ and Φ of each model and scenario parameter estimators ϕ and Φ.
11. Calculates the difference between the parameter and the mean of the ϕ and Φ parameter estimators.
12. The value of the difference is then used to see the interaction of the three factors.
13. Seeing the level of consistency of model accuracy from each model and scenario.
14. Exploration to identify large chili price data patterns by looking at data plots.
15. Divide the data into two parts, namely *training* data and *testing data.*
16. Checking for stationarity in mean and variance
17. Performing SARIMA modeling on *training data.*
18. Evaluate the model with data *testing.*
19. Comparing the RMSE and MAPE values of the SARIMA model.
20. Forecasting several future periods by adjusting the forecasting results.

3. RESULTS AND DISCUSSION

This section consists of simulation results for forecasting seasonal time series data with several types of data characteristics that have been determined by scenarios using the SARIMA method. Furthermore, the SARIMA method was applied to the actual data, namely the price of large chilies in five provinces on the island of Java.

3.1. Simulation Results

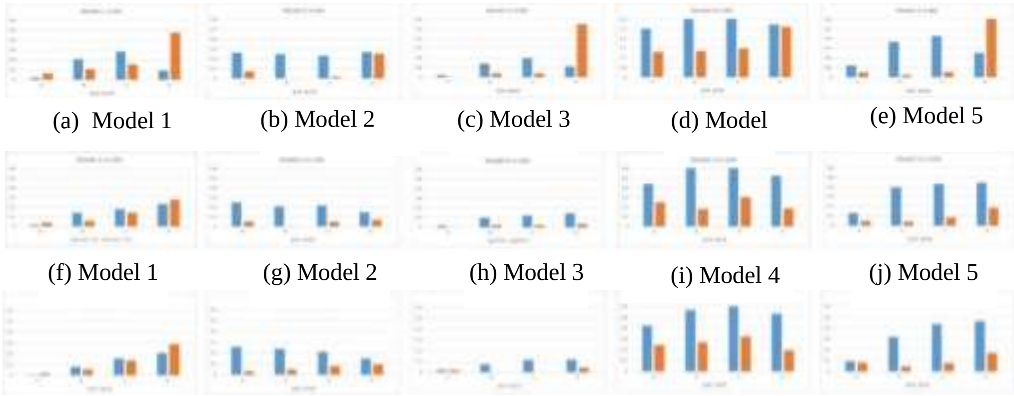
Based on the simulation for 100 repetitions, 100 estimators and parameters are obtained, then calculated on average. The mean of parameter estimators for and is presented in **Table 1**.

Table 1. Mean Estimator of $\hat{\phi}$ and $\hat{\Phi}$ Parameters

Model	Scenario	n	ϕ	Φ	$\lvert\hat{\phi}_{mean}\rvert$	$\lvert\hat{\Phi}_{mean}\rvert$
Model 1	Scenario A	84	0.63	0.29	0.61	0.36
		204	0.63	0.29	0.62	0.33
		504	0.63	0.29	0.62	0.31
⋮	⋮	⋮	⋮	⋮	⋮	⋮
Model 5	Scenario C	84	-0.72	-0.16	-0.29	-0.11
		204	-0.72	-0.16	-0.29	-0.08
		504	-0.72	-0.16	-0.28	-0.08
Model 5	Scenario D	84	-0.72	-0.16	-0.47	0.70

Model	Scenario	n	ϕ	Φ	$\lvert\hat{\phi}_{mean}\rvert$	$\lvert\hat{\Phi}_{mean}\rvert$
		204	-0.72	-0.16	-0.27	0.03
		504	-0.72	-0.16	-0.26	0.01

Table 1 shows the average value of 100 $\hat{\phi}$ and $\hat{\Phi}$ values generated from the generation and modeling process. These values are then subtracted from the actual parameter values to determine the magnitude of the difference or bias of the parameter estimators. Parameter estimation expects the estimated value to get closer to the actual value. The difference between values with parameters is presented in **Figure 1**.



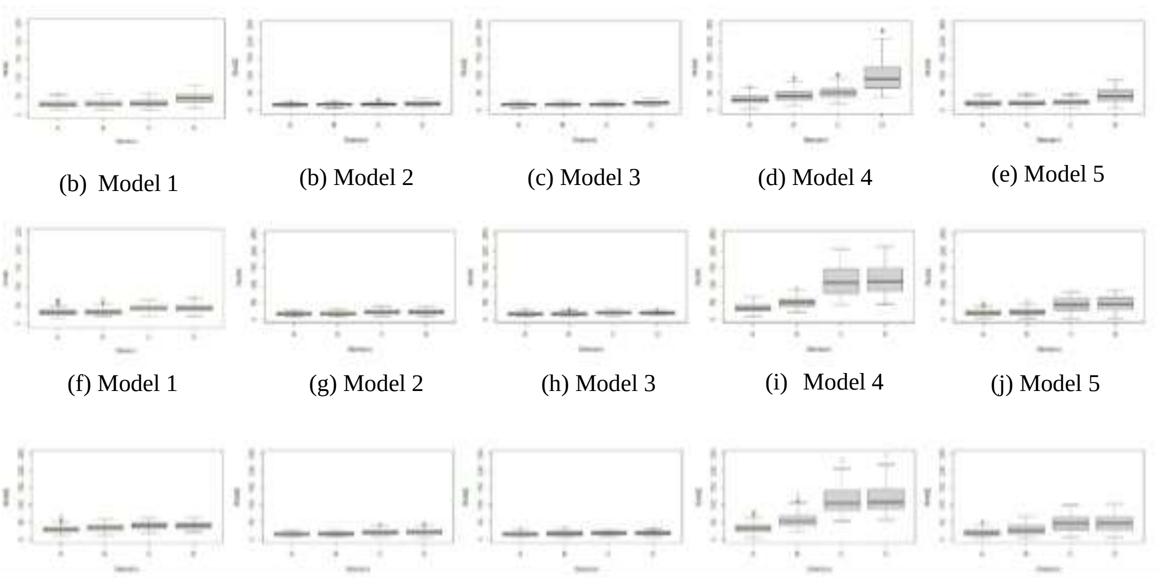
(a) Model 1 (b) Model 2 (c) Model 3 (d) Model (e) Model 5

(f) Model 1 (g) Model 2 (h) Model 3 (i) Model 4 (j) Model 5

(k) Model 1 (l) Model 2 (m) Model 3 (n) Model 4 (o) Model 5

Figure 1. Barplot of the mean difference between psi and mu parameter estimators with parameter (a) to (e) n = 84, (f) to (j) n = 204, (k) to (o) n = 504

Figure 1 shows the difference between the values of $\hat{\phi}$ and $\hat{\Phi}$ in each model and scenario. However, there is a tendency that if the number of outliers increases, the more significant the difference between the estimator value and the initial parameter is. Because of the large number of values, it can also be seen based on the RMSE value of the forecasting results. The distribution of the obtained RMSE values is presented in **Figure 2**.



(b) Model 1 (b) Model 2 (c) Model 3 (d) Model 4 (e) Model 5

(f) Model 1 (g) Model 2 (h) Model 3 (i) Model 4 (j) Model 5

(k) Model 1 (l) Model 2 (m) Model 3 (n) Model 4 (o) Model 5

Figure 2 Boxplot RMSE (a) to (e) Boxplot RMSE for n = 84, (f) to (j) Boxplot RMSE for n = 204, (k) to (o) Boxplot RMSE for n = 504

Figure 2 shows that the RMSE values in scenarios A, B, and C tend to be small in models 1, 2, and 3, while in models 4 and 5, the RMSE values in scenarios C and D tend to be larger. Based on this, it can be seen that the parameter value influences the consistency of the accuracy value when there is an outlier value. In addition, based on models 4 and 5, it can be seen that there is an influence from the scenario related to the number of outliers. The larger the number of outliers, the higher the RMSE value. The magnitude of the difference in the RMSE value for each sample size does not show a significant difference. This provisional conclusion can be proven by looking at the analysis of the variance table for each parameter. The following table presents the experimental design for the three factors in **Table 2**.

Table 2 Three Factor Experimental Designs for Estimating Parameters $\hat{\phi}$ and $\hat{\Phi}$

Factor A	Factor B	Factor C	Repetition	$\lvert\hat{\phi}_{mean} - \phi\rvert$	$\lvert\hat{\Phi}_{mean} - \Phi\rvert$
Model 1	Scenario A	84	1	0.02	0.07
Model 1	Scenario A	84	2	0.02	0.02
Model 1	Scenario A	84	3	0.03	0.04
Model 1	Scenario A	84	4	0.01	0.05
Model 1	Scenario A	84	5	0.01	0.03
⋮	⋮	⋮	⋮	⋮	⋮
Model 5	Scenario D	504	15	0.49	0.18
Model 5	Scenario D	504	16	0.50	0.18
Model 5	Scenario D	504	17	0.50	0.16
Model 5	Scenario D	504	18	0.48	0.18
Model 5	Scenario D	504	19	0.48	0.17
Model 5	Scenario D	504	20	0.49	0.18

Table 2 shows that the experimental design is to see the interaction of the three factors, each of which provides treatment according to its characteristics. The treatment results were measured based on the value of the parameter bias in each parameter. The following is presented in **Table 2** the analysis of variance from $\hat{\phi}$.

Table 3 Analysis of Variance $\hat{\phi}$

Factor	Degree of Freedom	Sum of Squares	Mean Square	Fhit	Ftab	Sig.
A	4	32.24	8.06	2652.23	2.38	0.00
B	3	4.37	1.46	479.83	2.61	0.00
C	2	0.25	0.12	40.48	3.00	0.00
AB	12	3.15	0.26	86.36	1.76	0.00
AC	8	0.35	0.04	14.40	1.95	0.00
BC	6	0.50	0.08	27.26	2.11	0.00
ABC	24	0.79	0.03	10.86	1.53	0.00
Error	1140	3.46	0.00			
Total	1199	45.11				

Based on **Table 3**, it can be seen that the three factors directly affect the magnitude of the parameter bias. Factor A, or parameter size, gives an F value of 2652.23 and is significant at 0.05. It shows a difference in the estimated parameter differences between models 1, 2, 3, 4, and 5, so the parameter size influences the difference in parameter estimates. Factor B, or outlier variation, gives an F value of 479.83 and is significant at 0.05. It shows differences in the estimated parameter differences between scenarios A, B, C, and D, so outlier variations affect the difference in parameter estimates. The same is valid for factor C or the sample size factor.

The results of the interaction of factors A, B, and C are also shown in **Table 3**. The interaction between factors A and B gives an F value of 86.36, indicating a difference in the estimated parameter differences due to the interaction of the two factors. The same thing also happened to the interaction of factors A and B and B. The interaction results of these three factors give an F value of 10.86 and a significant 0.05. Based on this value, it can be seen that there are differences in parameter estimates as the interaction of factors A, B, and C. The interaction results of these three factors give an F value of 10.86 and a significant 0.05. Based on this value, it can be seen that there are differences in parameter estimates as the interaction of factors A, B, and C. The results of the experimental design are presented in **Table 4**.

Table 4 Analysis of Variance $\hat{\Phi}$

Factor	Degree of Freedom	Sum of Squares	Mean Square	Fhit	Ftab	Sig.
A	4	7.08	1.77	289.15	2.38	0.00
B	3	6.18	2.06	336.88	2.61	0.00
C	2	1.89	0.95	154.71	3.00	0.00
AB	12	2.84	0.24	38.71	1.76	0.00
AC	8	0.50	0.06	10.13	1.95	0.00
BC	6	4.78	0.80	130.11	2.11	0.00
ABC	24	1.98	0.08	13.45	1.53	0.00
Error	1140	6.97	0.01			
Total	1199	32.22				

Overall, the Φ variance analysis table provides results not far from the $\hat{\phi}$ variance analysis. Each factor and the interaction between two or three factors showed significant results. Based on the two tables of analysis of variance, there is sufficient evidence that there is an effect of parameter size, outlier variation, and sample size on the magnitude of parameter bias. Therefore, paying attention to these three factors is necessary when applying forecasting methods to actual data. Factor A can be watched out for by choosing the suitable parameter estimation method, and factor B can be watched out for by choosing a robust method against outliers. In contrast, factor C needs to be wary of choosing a large amount of data because a small amount of data will cause parameter bias. Based on the simulation results, the recommended sample size that can be used is a minimum of 84 data.

3.2. The Results of the Application of Large Chili Prices

3.2.1. Data Exploration

The movement of large chili prices in five provinces on the island of Java from year to year is quite volatile. The annual development of large chili prices from 2015 to 2021 in five provinces is shown in **Figure 3**.

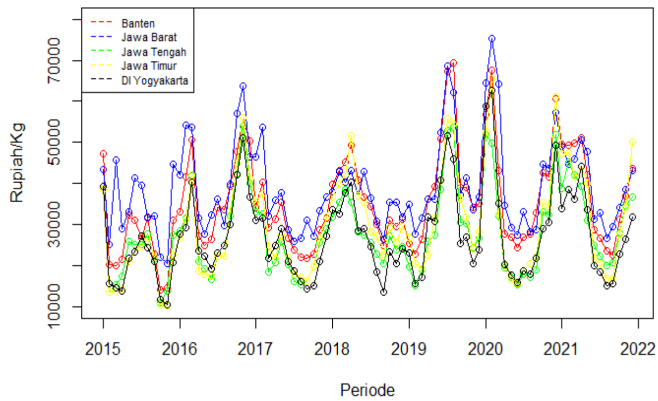


Figure 3. Graph of Large Chili Prices in Five Provinces

Based on **Figure 3**, prices in five provinces tend to have the same pattern in the last seven years. At the beginning of the year, prices tend to soar, and at the end of the year, prices tend to fall.

3.3.2. Stationary Test of Data

Stationary data is indicated by the constant mean and variance over time. **Figure 3** shows that chili price data in five provinces fluctuates around the median value with non-constant variances. Non-stationary data can also be seen from the formal test using the ADF test and the Bartlett test **[14]**, **[15]**. The results of non-seasonal and seasonal tests show that the data used are not stationary in variance. Here are the results of the stationarity test.

Table 5. Stationary Test Results in Variety and Average

Location	Non seasonal		Seasonal		Variety	
	p-value	Decision	*p-value*	Decision	*p-value*	Decision
Banten	0.01	Stationary	0.08	Not Stationary	0.0311	Not Stationary
Jawa Barat	0.02	Stationary	0.44	Not Stationary	0.0084	Not Stationary
Jawa Tengah	0.02	Stationary	0.03	Stationary	0.0212	Not Stationary
DIY	0.03	Stationary	0.01	Stationary	0.0362	Not Stationary
Jawa Timur	0.03	Stationary	0.17	Not Stationary	0.0291	Not Stationary
Location	Non seasonal		Seasonal		Variety	
	p-value	Decision	*p-value*	Decision	*p-value*	Decision
Banten	0.01	Stationary	0.08	Not Stationary	0.0311	Not Stationary
Jawa Barat	0.02	Stationary	0.44	Not Stationary	0.0084	Not Stationary
Jawa Tengah	0.02	Stationary	0.03	Stationary	0.0212	Not Stationary

Based on these results, transformation must be carried out to fulfill the assumptions. Data that has met the assumptions can be used to form candidate models. Model candidates were obtained based on ACF plots and PACF plots.

3.3.3. Estimating SARIMA Parameters

This stage aims to determine the feasibility of the parameters used in the model. The parameters of a model are said to be feasible if the t-count is greater than the t-table for all parameters. Based on the parameter estimation results, there are models with all significant parameter estimates and models with insignificant parameter estimates. Of course, the model chosen to be the best model is the model with all significant parameter estimates. If all parameters are significant, the model performance can be better.

3.3.4. Diagnostic Test of Residual Model SARIMA

The model that has been proven to have significant parameters and the smallest AIC value then examined the assumptions of freedom and normality of the residuals in the data using the Ljung-Box and Saphiro-Wilk tests [16]. The model is said to have residual p-values that are independent and normal if the value of the test results is greater than α = 0,05. The following is a model of each location, along with the results of the residual assumption test.

Table 6. The *p-value* of the Ljung-Box and Saphiro-Wilk tests

Locations	Model	Ljung-Box	Saphiro-Wilk
Banten	SARIMA(2,0,0)(2,1,0)[12]	0.373	0.066
Jawa Barat	SARIMA(1,0,0)(1,1,0)[12]	0.268	0.069
Jawa Tengah	SARIMA(1,0,0)(1,0,0)[12]	0.396	0.336
DIY	SARIMA(2,0,0)(1,1,0)[12]	0.322	0.064
Jawa Timur	SARIMA(2,0,0)(1,1,0)[12]	0.465	0.210

Next, overfitting to ensure the best model is selected. The results of overfitting show that no SARIMA model has a significant estimate of all parameters. Therefore, the best SARIMA model obtained from each location is listed in Table 6 because all the estimated parameter values are significant, there is no autocorrelation between the residuals, and the AIC value of the model is smaller than the other models. The next step is using the model as a reference for analyzing the SGSTAR method.

3.3.5. Model Evaluation and Forecasting Results

Forecasting is carried out for 12 periods, from January 2021 to December 2021. The forecasting results are then used to test the accuracy of the model. The accuracy level is tested by comparing the forecast results with testing data. The following is the forecast accuracy for each location by both methods.

Table 7. Evaluation of Models with MAPE Values

Locations	RMSE	MAPE
Banten	2.37	0.42
Jawa Barat	2.12	0.39
Jawa Tengah	3.06	-0.24
DIY	3.46	0.55
Jawa Timur	3.23	0.23

Table 7 shows that the RMSE and MAPE values tend to be small. It shows that the accuracy of the SARIMA model is good, and the model can be used to predict the price of large chili peppers in the following several periods. This conclusion aligns with the simulation results that the SARIMA method performs well if the data does not contain outlier values or a maximum of three outlier values at each location.

Next is forecasting the price of large chilies in Indonesia for the coming period, namely January-December 2022. The results of forecasting the prices of large chilies in Indonesia in the coming period are presented in **Figure 4.**

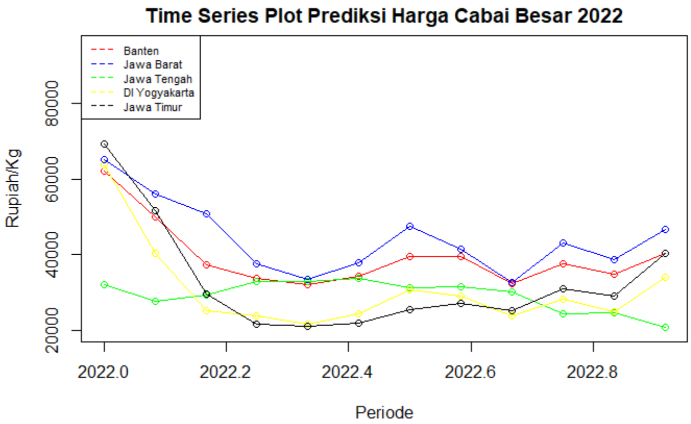


Figure 4. Plot of big chili price prediction 2022

Based on Figure 4, the price of large chilies in 2022 in the five provinces will still fluctuate. According to the forecast results for the next twelve periods, the price of big chili in five provinces is predicted to reach the highest price in the first three months of 2022. The highest price is predicted to occur in DIY Province in February, which is Rp. 74,230.00/kg. However, from the middle to the end of the year, prices will tend to fall and stabilize. The price will be the lowest in Middle Java Province in December, which is Rp. 20,689.00/Kg.

4. CONCLUSIONS

The simulation results on the generated data conclude that the parameter size, outlier value, and sample size can affect the forecast results. These three factors need to be considered when applying forecasting methods to actual data. Factor A can be watched out for by choosing the suitable parameter estimation method, and factor B can be watched out for by choosing a robust method against outliers. In contrast, factor C needs to be wary of choosing a large amount of data because a small amount of data will cause parameter bias. Applying the actual data using the SARIMA method gave good results. It can be seen based on the RMSE and MAPE values which tend to be small. Based on the forecasting results for the next 12 periods, it is predicted that the price of large chilies in 2022 in the five provinces will still fluctuate. Large chili prices in five provinces are predicted to reach the highest in the first three months of 2022. The highest price is predicted to occur in the DIY Province in February, which is Rp. 74,230.00/kg. However, from the middle to the end of the year, prices will tend to fall and stabilize. The price will be the lowest in Middle Java Province in December, which is Rp. 20,689.00/Kg.

REFERENCES

[1] BPS, *Pola Distribusi Perdagangan Komoditas Cabai Merah Tahun 2019*. 2019.

[2] D. C. Montgomery, C. L. Jennings, and M. Kulahci, “Introduction Time Series Analysis and Forecasting,” p. 671, 2015.

[3] L. Hakim, K. Kartomo, and N. Nursalam, “Dynamics of Price Growth of Curly Red Chili and Red Chili in Kolaka District,” *Agribus. J.*, no. December 2020, pp. 19–26, 2020, [Online]. Available: http://www.usnsj.com/index.php/AJ/article/view/1367

[4] K. Duangchaemkarn, W. Boonchieng, P. Wiwatanadate, and V. Chouvatut, “SARIMA Model Forecasting Performance of the COVID-19 Daily Statistics in Thailand during the Omicron Variant Epidemic,” *Healthc.*, vol. 10, no. 7, 2022, doi: 10.3390/healthcare10071310.

[5] N. Deretić, D. Stanimirović, M. Al Awadh, N. Vujanović, and A. Djukić, “SARIMA Modelling Approach for Forecasting of Traffic Accidents,” *Sustain.*, vol. 14, no. 8, 2022, doi: 10.3390/su14084403.

[6] N. P. N. Hendayanti and M. Nurhidayati, “Perbandingan Metode Seasonal Autoregressive Integrated Moving Average (SARIMA) dengan Support Vector Regression (SVR) dalam Memprediksi Jumlah Kunjungan Wisatawan Mancanegara ke Bali,” *J. Varian*, vol. 3, no. 2, pp. 149–162, 2020, doi: 10.30812/varian.v3i2.668.

[7] R. A. M. Yusuf and T. S. Yanti, “Perbandingan Metode Seasonal Autoregressive Integreted Moving Average (SARIMA) dan Metode Fuzzy Time Series untuk Model Peramalan Jumlah Wisatawan Mancanegara di Bali,” *Pros. Stat.*, pp. 597–605, 2021.

[8] S. H. Sørbye, P. G. Nicolau, and H. Rue, “Model-based bias correction for short AR(1) and AR(2) processes,” no. 1, pp. 1–21, 2020, [Online]. Available: http://arxiv.org/abs/2010.05870

[9] A. S. Ahmar *et al.*, “Modeling Data Containing Outliers using ARIMA Additive Outlier (ARIMA-AO),” *J. Phys. Conf. Ser.*, vol. 954, no. 1, 2018, doi: 10.1088/1742-6596/954/1/012010.

[10] E. B. Ferreira, P. P. Cavalcanti, and D. A. Nogueira, “ExpDes: An R Package for ANOVA and Experimental Designs,” *Appl. Math.*, vol. 05, no. 19, pp. 2952–2958, 2014, doi: 10.4236/am.2014.519280.

[11] G. Brassington, Brassington, and Gary, “Mean absolute error and root mean square error: which is the better metric for assessing model performance?,” *Eguga*, vol. 19, p. 3574, 2017, [Online]. Available: https://ui.adsabs.harvard.edu/abs/2017EGUGA..19.3574B/abstract

[12] D. C. Montgomery, *Design and Analysis of Experiments Eighth Edition. Arizona State University*, vol. 2009, no. 2005. 2013.

[13] I. Chang, G. C. Tiao, and C. Chen, “Linked references are available on JSTOR for this article : Time Series Model Specification in the Presence of Outliers,” vol. 81, no. 393, pp. 132–141, 2016.

[14] E. Paparoditis and D. N. Politis, “The asymptotic size and power of the augmented Dickey–Fuller test for a unit root,” *Econom. Rev.*, vol. 37, no. 9, pp. 955–973, 2018, doi: 10.1080/00927872.2016.1178887.

[15] D. Sharma and B. M. G. Kibria, “On some test statistics for testing homogeneity of variances: a comparative study,” *J. Stat. Comput. Simul.*, vol. 83, no. 10, pp. 1944–1963, 2013, doi: 10.1080/00949655.2012.675336.

[16] E. González-Estrada and W. Cosmes, “Shapiro–Wilk test for skew normal distributions based on data transformations,” *J. Stat. Comput. Simul.*, vol. 89, no. 17, pp. 3258–3272, 2019, doi: 10.1080/00949655.2019.1658763.