Forecasting A Weekly Red Chilli Price in Bengkulu City Using Autoregressive Integrated Moving Average (ARIMA) and Singular Spectrum Analysis (SSA) Methods

Novi Putriasari[1*], Sigit Nugroho[1], Ramya Rachmawati[1], Winalia Agwil[1], and Yosep O Sitohang[2]

[1]Mathematics Department, The University of Bengkulu, Indonesia
[2]Statistics Indonesia, Branch Office of Bengkulu Province, Indonesia

* Corresponding Author. E-mail: noviputriasari12@gmail.com

Article Info	Abstract
	Red chili occupies a strategic position in the Indonesian economic structure because its use applies to almost all Indonesian dishes. Therefore, controlling the price of red chili is a necessity to maintain national economic stability. The purpose of this research is to forecast a red chili weekly price using ARIMA and SSA based on the weekly data of chili prices from January 2016 - December 2019 sourced from Statistics Indonseia (BPS) Branch Office of Bengkulu Province. The data have been analyzed using software R. Based on MAPE, ARIMA (2,1,2) provides the best forecasting with value 0.49% while SSA 10.64%.

1. INTRODUCTION

A red chili is one of the vegetable plants whose existence is needed by most people, where its use applies to all daily cooking recipes. Therefore, the price of red chili is one of the factors that can affect the value of inflation, although the price of red chili often fluctuates following certain moments throughout the year. Such price behavior can have a negative impact on the economy at the national and local levels in various regions. Therefore, every effort to understand the pattern of red chili price behavior is a must so that simulations of various red chili prices can be carried out in the future as the basis for formulating government policies in overcoming inflation. . The substance of understanding or finding chili price behavior patterns is to build a mathematical model based on past data which is generally referred to as time series data. Therefore, efforts to build time series data from the behavior of red chili prices are important to do. One of the models that can be applied is the ARIMA model because this model can predict time series data that has non-stationary characteristics.

The procedure for building a time series model requires three stages, namely model identification, model fitting, and model diagnosis [2]. In the process of forming this model, the basis used is the assumption that all time series data is controlled by models which are partitioned into several model classes in addition to other assumptions such as linearity, normality and stationarity. The breakthrough to solve these problems is the formation of a new method, namely the Singular Spectrum Analysis (SSA) method. With this method, the assumptions used above are no longer valid or in other words are ignored, and therefore the perspective on model class classification becomes invalid [4]. Based on what has been mentioned previously, it can lead to a hypothesis that the SSA and ARIMA methods are the latest methods that can be used to predict the characteristics of red chili prices in the future.

2. METHOD

In this study, the data used are red chili price data in Bengkulu City (January 2016 – December 2019) obtained from the BPS Bengkulu Province. The data are analyzed using ARIMA and SSA methods with the help of program

R. The two kinds of method family in which this study is concerned with are ARIMA Methods and SSA Methods. The following is the practical explanations regarding the two method family.

2.1 ARIMA

The general form of the ARIMA model is as follows:

$$\phi_p(B)(1 - B)^d Z_t = \theta_q(B)a_t,$$

where Z_t is the actual value of the weekly red chili price at the t-th time, ϕ_p is the p-order of the AR parameter, θ_q is the q-order of the MA parameter, a_t is the error at the t-time, B is the backshift operator, d is the order of differencing parameters [1].

2.2 SSA

This methods need five stages, those are embedding, singular value decomposition, grouping, diagonal averaging, and finally forcasting.

2.2.1 Embedding

At this stage, the existing time series data will be converted into a new multidimensional data series with L×K size. The data series is formed into a path matrix as follows:

$$\mathbf{X} = [\boldsymbol{x_1}: \cdots : \boldsymbol{x_K}] = \begin{bmatrix} x_1 & x_2 & \cdots & x_K \\ x_2 & x_3 & \cdots & x_{K+1} \\ \vdots & \vdots & \ddots & \vdots \\ x_L & x_{L+1} & \cdots & x_N \end{bmatrix},$$

where L is parameter window length, and $K = N - L + 1$

2.2.2 Singular Value Decomposition (SVD)

At this stage, the path matrix X is decomposed into a matrix Z, where the matrix $Z = XX^T$. For the matrix Z, the eigenvalues and their corresponding eigenvectors can be calculated. If the rank of matrix X is denoted by r, then $v_i = \frac{X^T u_i}{\sqrt{\lambda_i}}$ will be obtained so that the path matrix X will be:

$$\mathbf{X} = \mathbf{X_1} + \cdots + \mathbf{X_r} = \sqrt{\lambda_1}\boldsymbol{u_1}\boldsymbol{v_1^T} + \cdots + \sqrt{\lambda_r}\boldsymbol{u_r}\boldsymbol{v_r^T}$$

where λ_i is eigenvalue, $\boldsymbol{u_i}$ is eigenvector and v_i is principle component.

2.2.3 Grouping

At this stage, the data resulting from the expansion of the matrix X are classified into types of time series such as trends, seasonality, cyclical, and noise.

2.2.4 Diagonal Averaging

At this stage, the grouping results of previous stage are transformed into new time series data through the following formula[3]:

1. $y_k = \frac{1}{k}\sum_{m=1}^{k} y_{m,k-m+1}^*$ for $1 \le k \le L^*$

2. $y_k = \frac{1}{L^*}\sum_{m=1}^{L^*} y_{m,k-m+1}^*$ for $L^* \le k \le K^*$

3. $y_k = \frac{1}{N-K^*+1}\sum_{m=1}^{N-K^*+1} y_{m,k-m+1}^*$ for $K^* < k \le N$

where $L^* = \min(L, K)$; $K^* = \max(L, K)$ and $N = L + K + 1$

2.2.5 Forecasting SSA

A common forecasting used for SSA is the type of recurrent SSA with the following equation:

$$(r_{L-1}, r_{L-2}, \cdots, r_1)^T = \frac{1}{1-v^2} \sum_{i=1}^{L-1} \pi_i \boldsymbol{u}_i^{\bar{v}}$$

where $\boldsymbol{u} = (u_1, u_2, \cdots, u_{L-1}, u_L)^T$, $\boldsymbol{u}^{\bar{v}} = (u_1, u_2, \cdots, u_{L-1})^T$, $\pi_i = u_L$ is the last component of eigenvector $\boldsymbol{u}$, and $v^2 = \sum_{i=1}^{r} \pi_i^2$.

3. RESULTS AND DISCUSSION

Before forecasting, the weekly red chili price data are divided into 2 parts, namely training data and testing data. The size of training data is larger than the one of testing data and aims to build or train the model, while the testing data is smaller in size and aims to validate the forecasting results of the model obtained from the training data. In this study, the training data size is 204 while the testing data size is 4.

3.1 ARIMA Forecasting

To forecast through ARIMA model, four steps which need to take across are examining data stationery, predicting initial model parameter, testing parameter eligibility, and model forecasting.

3.1.1 Examining Data Stationery

In this case, it can be seen in Figure 1 that the red chili price data used has an average that is not constant, while the Dickey Fuller test shows that the p-value <0.05. So it can be concluded that the data is not stationary.
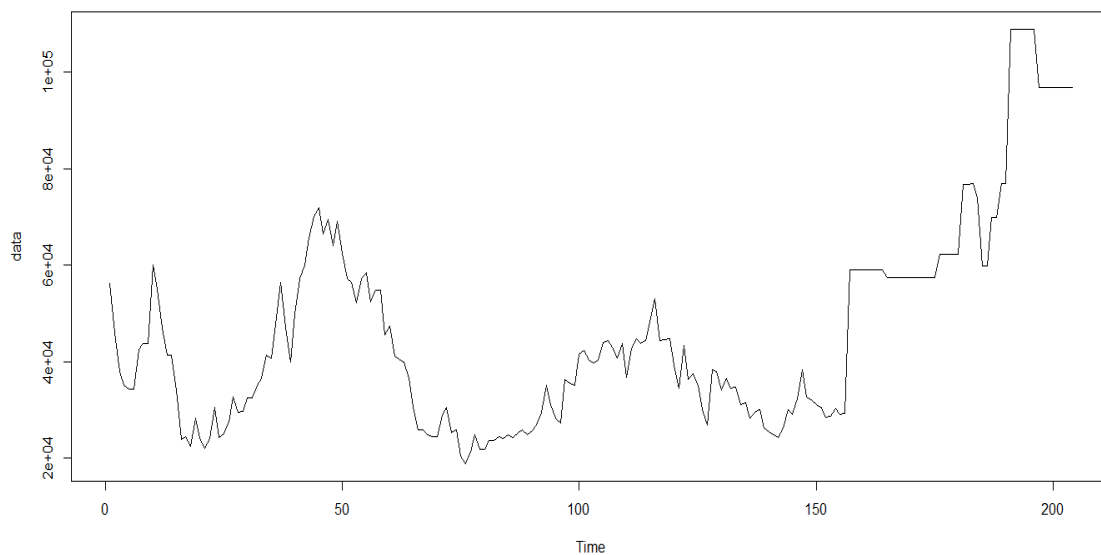

```
Augmented Dickey-Fuller Test

data:  data
Dickey-Fuller = -1.608, Lag order = 5, p-value = 0.7403
alternative hypothesis: stationary
```

Figure 2. Stationarity test results after differencing

3.1.2 Predicting Initial Model Parameter ARIMA (p, d, q)

At this stage, parameter estimation can be done by looking at the ACF and PACF plots. In Figure 3, it can be seen that the ACF plot looks cut-off at Lag 0, while the PACF plot is considered to be in exponential form, hence trial and error must be carried out using a range of values (0,5) for parameters p and q, while the value d= 1.
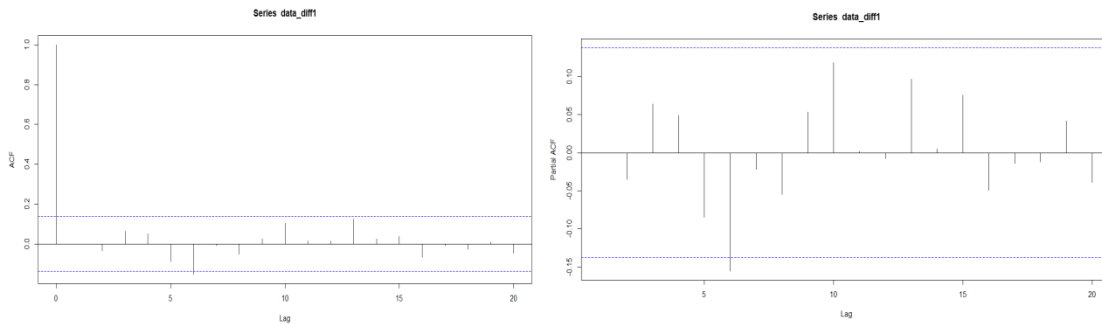
Figure 3. ACF and PACF plots

After trial and error, as shown by Figure 4 above, we get a model with all significant parameters and the probability less than 0.05, in which the only model is the model of ARIMA (2,1,2).

```
                                        z test of coefficients:

                                                Estimate Std. Error z value  Pr(>|z|)
Coefficients:                           ar1 -0.848757    0.024708 -34.352 < 2.2e-16 ***
        ar1      ar2     ma1     ma2     ar2 -0.958296    0.026705 -35.884 < 2.2e-16 ***
     -0.8488  -0.9583  0.9109  1.0000   ma1  0.910874    0.020063  45.400 < 2.2e-16 ***
s.e.  0.0247   0.0267  0.0201  0.0208   ma2  0.999985    0.020774  48.137 < 2.2e-16 ***
                                        ---
```

Figure 4. ARIMA trial and error results (2,1,2)

3.1.3 Testing Parameter Eligibility

At this stage, there are two tests carried out, those are the white noise test and the residual normality test. The following Figure 5 shows that all p-values are greater than $\alpha = 0.05$. It indicates that the residual data from the ARIMA model (2,1,2) meets the white noise requirements. In addition, it also shows that the residuals of the model ARIMA (2,1,2) have followed the line. Therefore it can be said that the residuals of the model ARIMA (2,1,2) are normally distributed.

```
 m    Qm      pvalue
 1   0.31  0.57688910
 2   0.33  0.56410317
 3   1.14  0.28502220
 4   1.15  0.28445938
 5   1.63  0.20145942
 6   5.78  0.05554667
 7   6.27  0.09913542
 8   6.43  0.16911808
 9   6.84  0.23306317
10   7.67  0.26346483
11   7.83  0.34750266
12   8.23  0.41152015
13   9.82  0.36526795
14   9.91  0.44833089
15  11.31  0.41794336
16  13.77  0.31568680
17  13.78  0.38951062
18  13.79  0.46524411
19  13.91  0.53233669
20  14.51  0.56097453
21  15.87  0.53320345
22  15.91  0.59855653
23  16.42  0.62878062
24  16.86  0.66231880
25  18.80  0.59826108
26  18.81  0.65682077
27  18.82  0.71144740
28  24.07  0.45747416
29  24.11  0.51312920
30  24.66  0.53848755
```
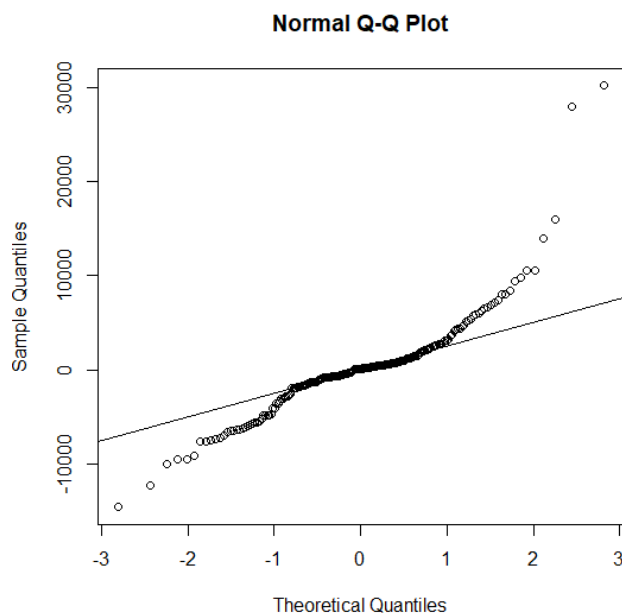

Figure 5. White noise test results and normality of ARIMA (2,1,2)

3.1.4 ARIMA (2,1,2) Model Forecasting

With the help of R program, the results are as in Table 1 with the MAPE value of 0.48694% and is included in the category of highly accurate forecasting. Hence forecasting can be carried out for the next 4 weeks as in Table 2.

Table 1. Forecasting Data Using ARIMA (2,1,2)

Week	Testing Data	Forecasting Data
205	96900	96103.88
206	96900	96583.39
207	96900	96939.32
208	96900	96177.71

Table 2. Price Forecasting Data Using ARIMA (2,1,2)

Week	Forecasting Data
209	96483.05
210	96953.74
211	96261.63
212	96398.00

## 3.2 SSA Forecasting

3.2.1 Embedding

In this process, the first thing to do is to determine the optimal window length (L) value, which is 46, so that an X matrix with size 46×159 is formed.

3.2.2 Singular Value Decomposition (SVD)

In the SVD process, the matrix X is decomposed into a matrix $\mathbf{Z} = \mathbf{X} \times \mathbf{X}^{\mathbf{T}}$ and then the rank of the X matrix is searched for $X_i, i = 1,2,3, \dots ,46$

3.2.3 Grouping

At this stage, grouping is done based on the plot of the eigenvectors on the Z matrix which can be seen in the images below. In this study, the number of eigenvector plots is 10 plots.
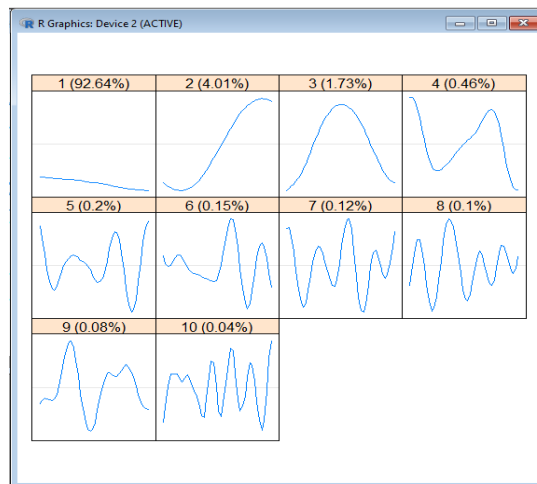
Figure 6. Plot of eigenvector for weekly chili price data

Based on Figure 6 above, the matrix $\mathbf{X_1}$ will be grouped into trend groups, while the matrices $\mathbf{X_2}$ and $\mathbf{X_3}$ appear to have oscillatory patterns, but the other matrices tend to have unclear patterns. Therefore in this case the matrices are classified into noise family.

3.2.4 Averaging Diagonal

At this stage, the group results of the grouping stage are transformed into new time series data of size N, so that a single value is obtained as the input data for the LRF forecasting process.

3.2.5 LRF (Linear Recurrent Formula)

The single value obtained of the previous process is the input for the LRF method. Forecasting data using LRF can be seen in Table 3.

Table 3. Chili Price Forecasting Using SSA

Week	Testing Data	Forecasting Data
205	96900	86814
206	96900	86759
207	96900	86576
208	96900	86269

4. CONCLUSION

The conclusion from this research is that the forecasting method that has the highest accuracy in predicting the price of red chili in Bengkulu City is the ARIMA (2,1,2) method with a MAPE of 0.48694 %, while the SSA method has a MAPE of 10.64218%.

REFERENCES

[1] Box, G.E.P., and Jenkins, G.M., Time Series Analysis: Forecasting and Control, Third Edition, New Jersey, Prentice Hall, 1994.
[2] Cryer, J.D., and Chan, K., Time Series Analysis with application in R, USA, Springer, 2008.
[3] Golyandina, N., Korobeynikov, A., and Zhigljavsky, A., Singular Spectrum Analysis With R, United Kingdom, Springer, 2018
[4] Hassani, H., and Mahmoudvand, R., Singular Spectrum Analysis Using R, Palgrave Macmillan, 2018