Forecasting Chilli Requirement with ARIMA Method

E Abinowi* and I D Sumitra

Postgraduate Faculty of Indonesian Computer University, Universitas Komputer Indonesia, Indonesia

*egi.abinowi@yahoo.com.sg

Abstract. The aim of the study Forecasting Chilli Requirement with ARIMA Method. The unbalanced over production causes the market price and production to be less than the amount of public consumption. Forecasting is the art and science of predicting events that will occur by using historical data and projecting it into the future with some form of mathematical modelling. To do the forecasting required a particular method and which method is used depends on the type of data to be predicted as well as the goal to be achieved. In this research using time series ARIMA forecasting method with data used data of chilli requirement and production data year 2011 until 2014. From resulted model hence calculated result of Mean Absolut Error (MAE) to see average of absolute value error at predicted result. The results of model trials conducted using ARIMA method (1, 1, 2) yield MAE of 12.18.

1. Introduction
In the past few years, the need for chilli commodities often results in an imbalance between crop production and the number of needs. It affects a certain period of high demand but the production of farmers' harvests cannot be sufficient, will have an effect on changing the market price, while in a certain time the amount of crop production exceeds that of the market price.

ARIMA also has limitations on accuracy of predictions yet it is used more widely for forecasting the future successive values in the time series [1]. Forecasting findings have shown, production increased on an annual basis [2]. The influence of forecasting period on the ARIMA model is found that the model accuracy decreases as the forecasting period increases [3]. The aim of the study Forecasting Chilli Requirement with ARIMA Method

2. Experimental Method
Forecasting is the art and science of predicting events that will occur by using historical data and projecting it into the future with some form of mathematical model [4]. Time series analysis is proposed as the usually used solution for time correlated data in statistics [5].

3. Result and Discussion
The data for this research is data of Bandung chili commodity consumption, and data of small chili production in 2010 – 2014 (See Table 1).
Table 1. Small Chili Consumption in 2011 – 2014.

| Month   | 2011    | 2012    | 2013    | 2014    |
|---------|---------|---------|---------|---------|
| January | 36.026  | 45.572  | 48.266  | 51.725  |
| February| 38.220  | 53.267  | 47.596  | 53.547  |
| March   | 44.777  | 70.360  | 56.367  | 68.900  |
| April   | 47.964  | 56.672  | 70.718  | 85.159  |
| Mei     | 56.285  | 62.653  | 65.522  | 78.219  |
| June    | 52.684  | 62.653  | 65.522  | 78.219  |
| July    | 60.029  | 69.700  | 70.718  | 75.514  |
| August  | 56.285  | 62.653  | 65.522  | 78.219  |
| September| 52.684 | 59.799  | 61.165  | 66.048  |
| October | 52.294  | 52.025  | 62.445  | 61.933  |
| November| 45.315  | 51.754  | 62.847  | 55.775  |
| December| 43.960  | 44.053  | 56.165  | 63.827  |

Generally speaking, ARIMA (p,d,q) class class consists of AR(p), MA(q), and ARMA(p, q) classes. Box and Jenkins proposed a general ARIMA model to cope with the modeling of non-stationary time series [5]. AR and MA process ARIMA models require stationary data [6]. In making the ARIMA model the main requirement is the stationary data, on average or in distance. The data can be said to be stationary if the data fluctuation is around a constant value (stationary in averages) and the range of fluctuations remains constant over time (stationary in variety). ARIMA(p; d; q) is combination of AR(p), MA(q) dan ARMA(p,q) classes [7].

In the AR model the current value of the variable is defined as a function of its previous values plus an error term. Mathematically, it is written as,

\[ x_t = \alpha_1 x_{t-1} + \alpha_2 x_{t-2} + \cdots + \alpha_p x_{t-p} + \epsilon_t \] (1)

The moving average is the average of the actual observations, t , whereas in this case it is a function of the error terms.

\[ x_t = \epsilon_t + \beta_1 \epsilon_{t-1} + \beta_2 \epsilon_{t-2} + \cdots + \beta_q \epsilon_{t-q} \] (2)

In the autoregressive and moving average models the Box-Jenkins method is known as the ARMA model with a stationary assumption and the ARMA model is written as,

\[ x_t = \alpha_1 x_{t-1} + \cdots + \alpha_p x_{t-p} + \epsilon_t + \beta_1 \epsilon_{t-1} + \beta_2 \epsilon_{t-2} + \cdots + \beta_q \epsilon_{t-q} \] (3)

The parameters p and q represents autoregressive and moving average respectively [8]. The time series needs to be differentiated before applying ARMA(p; q) model. ARIMA includes the differentiating operator d [8].

- **Model Identiﬁcation**: In the Model identiﬁcation phase the d value has to be set. It decides the stationary (d D 0) or non-stationary (d > 0) behavior of a time series. ACF and PACF plots are plotted to _nd out the parameters. The identiﬁcation of (p, q) is based on Akaike Information Criterion (AIC). The model with smallest AIC is chosen [8].
- **Estimation**: In this phase, the coefﬁcient _p and _q are estimated [8].
- **Diagnostic Checking**: The diagnostic phase deals with model adequacy by plotting the residuals. The model with the smallest residual is chosen [8].

From the data request of Bandung chili from January 2011 to 2013 processed using minitab statistical software package to produce as follows. And data of 2014 as comparison data for forecasting results.
Figure 1. Autocorrelation Function (ACF).

Figure 2. Partial Autocorrelation Function (PACF).

The ACF and PACF Figure 1 and Figure 2 are the result of the first difference, from the graph showing that nothing comes out of the line to show that the difference is 1 or D(1). Then the forecasting result as follows. Auto correlation function indicated the order of the autoregressive components ‘q’ of the model, while the partial correlation function gave an indication for the parameter p [9]. (See Table 2,3).

Table 2. Final Estimates of Parameters (2,1,1).

| Type  | P     |
|-------|-------|
| AR 1  | 0.000 |
| MA 1  | 0.000 |
| MA 2  | 0.234 |
Table 3. Forecasting’s from period 36.

| Period | Forecasting |
|--------|-------------|
| 37     | 54,9940     |
| 38     | 55,0984     |
| 39     | 55,1841     |
| 40     | 55,2546     |
| 41     | 55,3124     |
| 42     | 55,3600     |
| 43     | 55,3991     |
| 44     | 55,4312     |
| 45     | 55,4575     |
| 46     | 55,4792     |
| 47     | 55,4970     |
| 48     | 55,5116     |

From the results of the model test from Table 2, and Table 3 on ARIMA (1,1,2), there are AR and MA (1) which have P value from table 2 less than 0.050. Berk et al [10] P-value is determined by the software as 0.05 a-level corresponding 95% of confidence interval. If the P-value is less than this value, H0 is rejected. In [11] Calculation of MAE is relatively simple. It involves summing the magnitudes (absolute values) of the errors to obtain the ‘total error’ and then dividing the total error by n. Thus obtained MAE value is 12.18 obtained from the original data in 2014 and forecasting results in 2014. From the test results of chili demand can be supplied by the supply of peppers in accordance with forecasting in a manner similar to forecasting data supply of chili in 2011 to 2013 to get forecasting in 2014.

4. Conclusions

From the research process with ARIMA method on chilli demand data of Bandung City from January 2011 to December 2013 period of forecast forecasting test can be done using ARIMA. With ARIMA (1,1,2) yield MAE 12.18. By knowing pepper forecasting, it can help to meet the needs.

References

[1] M Kumar and M Anand 2014 An application of time series ARIMA forecasting model for predicting sugarcane production in India Studies in Business and Economics 9(1) pp. 81-94
[2] E A Frah 2016 Sudan Production of Sorghum; Forecasting 2016-2030 Using Autoregressive Integrated Moving Average ARIMA Model American Journal of Mathematics and Statistics 6 4 pp. 175-181
[3] Z Yu, G Lei, Z Jiang and F Liu 2017 ARIMA modelling and forecasting of water level in the middle reach of the Yangtze River Transportation Information and Safety (ICTIS), 2017 4th International Conference Banff, Canada
[4] D C Montgomery, L A Johnson and J S Gardiner 1990 Forecasting and Time Series Analysis McGraw-Hill Companies
[5] G E P Box and G Jenkis 1994 Time Series Analysis: Forecasting and Control (New York: Prentice-Hall) Holden-Day 3rd Edition
[6] K Dmytro, T Sergii and P Andiy 2017 ARIMA forecast models for scheduling usage of resources in IT-infrastructure Computer Sciences and Information Technologies (CSIT), 2017 12th International Scientific and Technical Conference
[7] W Wu, W Zhang, Y Yang and Q Wang 2010 Time series analysis for bug number prediction \textit{Engineering and Data Mining (SEDM), 2010 2nd International Conference}

[8] J Pati, B Kumar, D Manjhi and K K Shukla 2017 A Comparison Among ARIMA, BP-NN, and MOGA-NN for Software Clone Evolution Prediction \textit{IEEE Access 5} pp. 11841-11851

[9] N Iqbal, K Bakhsh, A Maqbool and A S Ahmad 2005 Use of the ARIMA model for forecasting wheat area and production in Pakistan \textit{Journal of Agriculture and Social Sciences 1(2)} pp. 120-122

[10] I Berk and V Ş Ediger 2006 Forecasting the coal production : Hubbert curve application on Turkey's lignite fields \textit{Resources Policy 50} pp. 193-203

[11] C. J. Willmott and K. Matsuura, “Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance,” \textit{Climate research 30 no. 1}, pp. 79-82, 2005.