International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 10 Issue 03, March-2021

Time Series Modeling and Forecasting for Indonesian Coffee Export

Zul Amry
Department of Mathematics,
State University of Medan, Indonesia

***Abstract*—This paper to build a time series forecasting model for data of Indonesian coffee export from 1976 until 2019. The method used in this research is the Box Jenkins method. The autocorre-
lation function (ACF) and the partial autocorrelation function (PACF) are used for stationary test and model idenfication. Ljunc Box Q statistics are used for diagnostic test, whereas to show the accuracy model are used Root Mean Squared Error (RMSE), Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE).**

Keywords*—*ARIMA model, coffee export, forecasting, Box Jenkins method.

I. INTRODUCTION

Indonesia is one of the largest coffee exporting countries in the world and the demand for coffee exports to increase every year. In order for this demand to be fulfilled, it is necessary to forecast the demand coffee in the future. One of the tools for forecasting is the time series model forecasting. Time series model forecasting is a type of forecasting that uses past observational data, investigates its behavior and is extrapolated into the future. The Autoregressive Integrated Moving Average (ARIMA) model developed by Box and Jenkins (1976) and has been widely used in various fields as a statistical model, especially related to forecasting problems. In connection with the brief description above, this paper focuses on constructing a time series forecasting model with the ARIMA model to be applied to Indonesian coffee export

II. MATERIALS AND METHODS

The material used in this paper consists of coffee export data and the theories of statistical related to forecasting time series models. Coffee export data is annual data of Indonesia coffee export from 1975 to 2019. The method used is the Box-Jenkins method. Some statistical theories in time series analysis are ARIMA model, ACF, PACF, maximum likelihood method, Ljung-Box Q statistics, RMSE, MAE and MAPE.

The *ARIMA (p, d, q)* model of the time series $\{x_1, x_2, \cdots\}$ is defined as

$$\Phi_p(B)\,\Delta^d x_t = \Theta_q(B)\epsilon_t \quad (1)$$

where B is the backward shift operator, $Bx_t = x_{t-1}$, $\Delta = 1 - B$ is the backward difference, $\Phi_p = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p$, $\Theta_q = 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q$

Principle of maximum likelihood yields a choice of the estimator as the value for the parameter that maximizes the likelihood function. If $X = (X_1, X_2, \cdots, X_n)$ represents a random sample from $f(x;\theta)$,then the likelihood function is:

$$L(\theta) = \prod_{i=1}^{n} f(x_i;\theta) \quad (2)$$

The maximum likelihood estimator, that is $\hat{\theta}$, is a value of θ that satisfies:

$$f(x_1, x_2, \cdots, x_n;\, \hat{\theta}) = f(x_1, x_2, \cdots, x_n;\, \theta) \quad (3)$$

The autocorrelation between Y_t and Y_{t+k} is defined as

$$\rho_k = \frac{Cov(Y_t, Y_{t+k})}{\sqrt{Var(Y_t)}\,\sqrt{Var(Y_{t+k})}} \quad (4)$$

that can be estimated from sample data by

$$\hat{\rho}_k = \frac{\hat{\gamma}_k}{\hat{\gamma}_0} = \frac{\sum_{t=1}^{n-k}(Y_t - \bar{Y})(Y_{t+k} - \bar{Y})}{\sum_{t=1}^{n}(Y_t - \bar{Y})^2} \quad (5)$$

and the set $\{\hat{\rho}_k, k = 0, 1, 2, \ldots\}$ is called the ACF.

The partial autocorrelation between Y_t and Y_{t+k} is defined as

$$\phi_{kk} = \frac{Cov\left[(Y_t - \hat{Y}_t), (Y_{t+k} - \hat{Y}_{t+k})\right]}{\sqrt{Var\left(Y_t - \hat{Y}_t\right)}\sqrt{Var\left(Y_{t+k} - \hat{Y}_{t+k}\right)}} \quad (6)$$

where $\hat{Y}_{t+k} = \alpha_1 Y_{t+k-1} + \alpha_2 Y_{t+k-2} + \cdots + \alpha_{k-1} Y_{t+1}$ and ϕ_{kk} can be estimated from sample data by

$$\hat{\phi}_{kk} = \frac{\begin{vmatrix} 1 & \hat{\rho}_1 & \hat{\rho}_2 & \cdots & \hat{\rho}_{k-2} & \hat{\rho}_1 \\ \hat{\rho}_1 & 1 & \hat{\rho}_1 & \cdots & \hat{\rho}_{k-3} & \hat{\rho}_2 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ \hat{\rho}_{k-1} & \hat{\rho}_{k-2} & \hat{\rho}_{k-3} & \cdots & \hat{\rho}_1 & \hat{\rho}_k \end{vmatrix}}{\begin{vmatrix} 1 & \hat{\rho}_1 & \hat{\rho}_2 & \cdots & \hat{\rho}_{k-2} & \hat{\rho}_{k-1} \\ \hat{\rho}_1 & 1 & \hat{\rho}_1 & \cdots & \hat{\rho}_{k-3} & \hat{\rho}_{k-2} \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ \hat{\rho}_{k-1} & \hat{\rho}_{k-2} & \hat{\rho}_{k-3} & \cdots & \hat{\rho}_1 & 1 \end{vmatrix}} \quad (7)$$

and the set $\{\hat{\phi}_{kk}, k = 0, 1, 2, \ldots\}$ is called the PACF.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 10 Issue 03, March-2021

The ACF and PACF use to identify the models. The following **Table 1** summarizes how to identify the model of the stationary data by using of characteristics for the ACF and PACF.

Table 1: Characteristics for the ACF and PACF

Model	ACF, ρ_k	PACF, ϕ_{kk}
AR(*p*)	Damped exponential and / or sine functions	ϕ_{kk} =0 for k>p
MA(*q*)	ρ_k = 0 for k >q	Dominated by damped exponential and/or sine function
ARMA(*p,q*)	Damped exponential and/or sine functions after lag (q-p)	Dominated by damped exponential and/or sine function after lag (p-q)

Diagnostics checking aims to conclude whether the forecasting model which obtained is adequate. the way is to test the assumption of residual independence between lags. If the residual is whit noise, then the model is adequate. The hypothesis is $H_0: \rho_1 = \cdots = \rho_k = 0$ vs $H_1: \exists_j, \rho_j \neq 0$ and tested with Ljung-Box Q Statistic

$$Q = n(n-2)\sum_{k=1}^{K} \frac{\hat{\rho}_k^2}{(n-k)} \sim \chi^2(K-p-q) \qquad (8)$$

where n is the sample size, $\hat{\rho}_k^2$ is the autocorrelation of residuals at lag k and K is the number of lags being tested, and reject H_0 at the level α, if $Q > \chi^2_{1-\alpha}\ (K-p-q)$.

The measures to determine the accuracy of a forecasting model in this research is RMSE, MAE and MAPE defined respectively as follows:

$$RMSE = \sqrt{\frac{ESS}{n}} \qquad (9)$$

$$MAE = \frac{\sum_{t=1}^{n} \left| Y_t - \hat{Y}_t \right|}{n} \qquad (10)$$

$$MAPE = \frac{\sum_{t=1}^{n} \left| \frac{Y_t - \hat{Y}_t}{Y_t} \right|}{n} \qquad (11)$$

where Y_t =The actual value at time t ; $\hat{Y}_t$ = The forecast value at time t; n =The number of observations and *ESS*=the error sum of square.

III. CONTRUCTION OF FORECASTING MODEL

The graph of the x original data in **Figure 1** shows an increasing trend and the ACF plot in **Figure 2** shows a slow decline, this indicates that the time series data is not stationary.

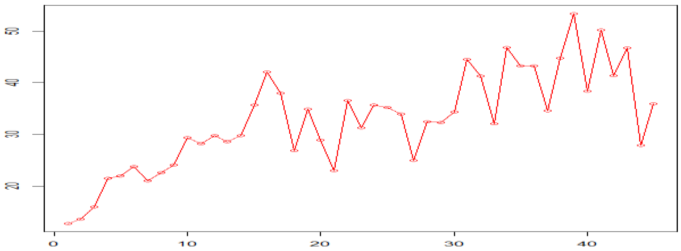

Figure 1: Plot of x

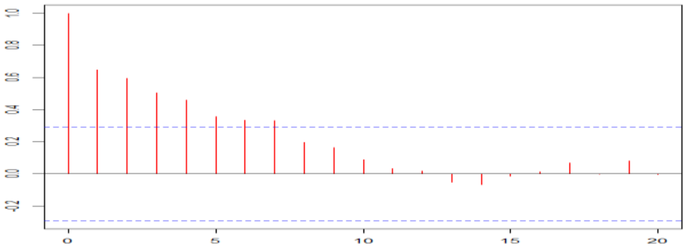

Figure 2: Plot ACF of x

To overcome this condition, it is transformed to y ; y is the first differences of x, that is $y_t = x_t - x_{t-1}$. The graph of the y data in **Figure 3** below and the ACF plot in **Figure 4** with a muffled sine wave shape, indicates that the time series data is stationary.

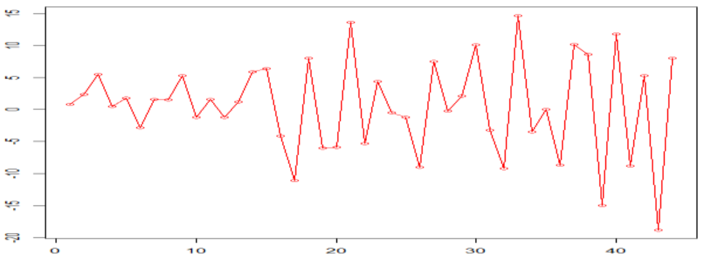

Figure 3: Plot of y

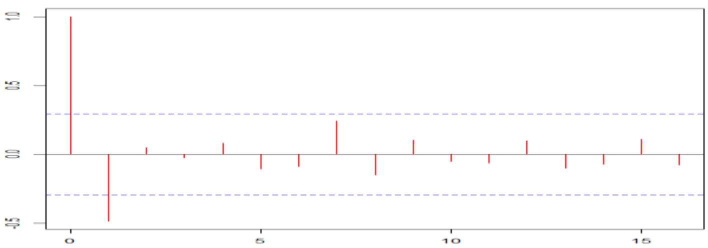

Figure 4: Plot ACF o y

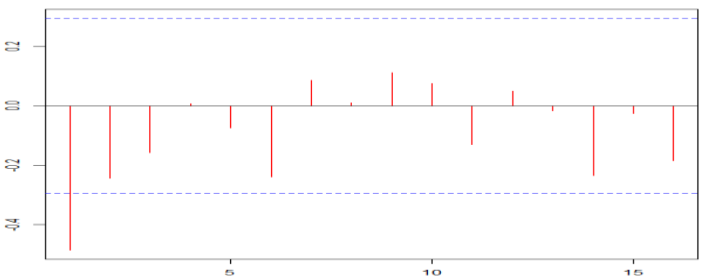

Figure 5: Plot PACF of y

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 10 Issue 03, March-2021

Furthermore, the construction of forecasting model consists of identification of model, estimation of parameter, diagnostic test, and accuracy of model. Based the ACF plot in **Figure 4** and the PACF plot in **Figure 5**, they are interrupted after lag 1, then the possible models for y data are ARMA (1, 0), ARMA (0, 1) or ARMA (1,1) or ARIMA (1,1,0), ARMA (0,1,1) or ARIMA (1, 1, 1) for x data. Results of estimation of parameters use likelihood maximum method presented in the **Table 2** below:

Tabel 2: Estimation of parameter

Model	estimate value of parameters	
	$\hat{\phi}_1$	$\hat{\theta}$
ARIMA (1,1,0)	-0.3433	-
ARMA (0,1,1)	-	-0.4693
ARIMA (1,1,1)	0.0588	-0.5131

Results of diagnostic test use Ljunc-Box Q statistics to ARIMA (1,1,0), ARMA (0,1,1) and ARMA (1, 1, 1) at the level $\alpha = 0.05$ and degrees of freedom=15 with the value $\chi^2_{0.95}\,(15) = 24.996$ presented in the Figure 6, Figure 7, Figure 8, and Table 3 below:

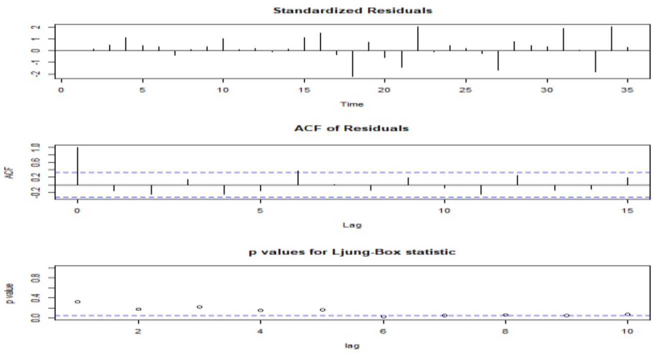

Figure 6: Diagnostic test of ARIMA (1,1,0) model

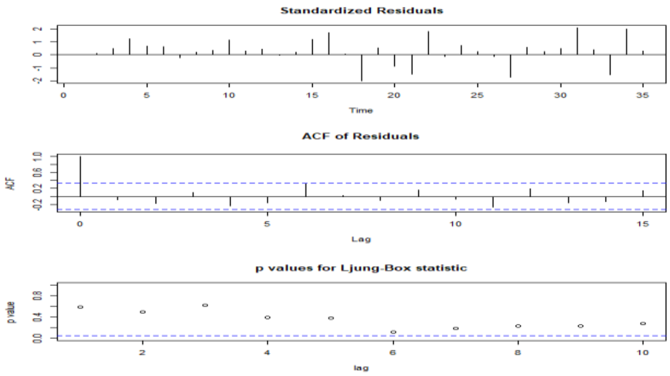

Figure 7: Diagnostic test of ARIMA (0,1,1) model

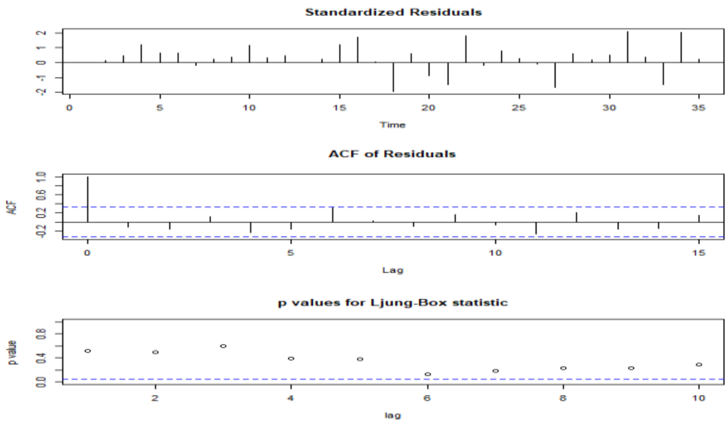

Figure 8: Diagnostic test of ARIMA (1,1,1) model

Tabel 3: Diagnostic test

Model	Q-value	Decision for H_0
ARIMA (1,1,0)	20.7242	No reject
ARMA (0,1,1)	15.5253	No reject
ARIMA (1,1,1)	15.7478	No reject

the results of diagnostic test in the **Table 3** above concluded that all models were adequate.

Results the accuracy of the model use RMSE, MAE and MAPE presented in the Table 4 below:

Tabel 4: Accuracy; RMSE, MAE and MAPE

Model	Accuracy		
	RMSE	MAE	MAPE
ARIMA (1,1,0)	7.7129	6.2174	16.9425
ARMA (0,1,1)	7.3466	6.0200	15.8802
ARIMA (1,1,1)	7.3236	6.0147	15.7893

Next, based on the smallest values of RMSE, MAE and MAPE in the **Table 4** above, it is concluded that the most suitable model is the ARIMA (1,1,1) model, that is

$$\Phi_1(B)\,\Delta^1 x_t = \Theta_1(B)\epsilon_t$$

$$\Leftrightarrow \quad (1-\phi_1 B)\,(1-B)x_t = (1-\theta_1 B)\epsilon_t$$

$$\Leftrightarrow \quad (1-\phi_1 B)\,(x_t - x_{t-1}) = \epsilon_t - \theta_1 B\epsilon_t$$

$$\Leftrightarrow \quad x_t - x_{t-1} - \phi_1 B x_t + \phi_1 B x_{t-1} = \epsilon_t - \theta_1 B\epsilon_t$$

$$\Leftrightarrow \quad x_t - x_{t-1} - \phi_1 x_{t-1} + \phi_1 x_{t-2} = \epsilon_t - \theta_1 \epsilon_{t-1}$$

$$\Leftrightarrow \quad x_t - 1.0588\,x_{t-1} + 0.0588\,x_{t-2} = \epsilon_t + 0.5311\epsilon_{t-1}$$

$$\Leftrightarrow \quad x_t = 1.0588\,x_{t-1} - 0.0588\,x_{t-2} + 0.5311\epsilon_{t-1} + \epsilon_t$$

IV. CONCLUSION

Time series data of Indonesian coffee export from 1976 until 2019 is not stationary, but stationary for one level difference data, so that the data analyzed is the difference of one level and the results are returned to the original data. Based on calculations and analysis of data it is concluded that the most suitable model is the ARIMA (1,1,1).

REFERENCES

[1] Alfaki, M. M and Masih, S. B (2015). Modeling and Forecasting by using Time Series ARIMA Models. *International Journal of Engineering Research & Technology (IJERT), Vol. 4 Issue 03, 914-918.*http://dx.doi.org/10.17577/IJERTV4IS030817

[2] Bain, L. J., & Engelhardt, M. (1992). *Introduction to Probability and Mathematical Statistics* (2nd ed.). Duxbury Press, Belmont, California. https://doi.org/10.2307/2532587

[3] Box, G.E.P. & Jenkins, G.M. (1976) Time Series Analysis: Forecasting and Control. Holden-Day, San Francisco.

[4] BPS-Statistics Indonesia (2020), *Indonesian Coffee Statistics 2019.*

[5] Ihaka, R. (2005). *Time Series Analysis*. Statistics Department University of Auckland.

[6] Directorate General of Estate Crops (2019). *Tree Crop Estate Statistics of Indonesia 2018- 2019,* Jakarta.

[7] Fattah, J. et. al. (2018), Forecasting of demand using ARIMA. *International Journal of Engineering Business Management, Vol.10, 1-9,* https://doi.org/10.1177/1847979018808673

[8] Jain, G. and Mallick, B. (2017), A Study of Time Series Models ARIMA and ETS, *I.J. Modern Education and Computer Science, 4, 57-63.* (http://www.mecs-press.org/) DOI: 10.5815/ijmecs.2017.04.07

[9] Madsen, H. (2008). *Time Series Analysis*. Chapmann Hall, Informatics and Mathematical Modelling, Technical University of Denmark. https://doi.org/10.1201/9781420059687

[10] Mgaya, J. F (2019). Application of ARIMA models in forecasting livestock products consumption in Tanzania *Cogent Food & Agriculture Vol. 5, Issue 1, 1-29,* https://doi.org/10.1080/23311932.2019.1607430

[11] Rossiter, D. G. (2013). *Time Analysis in R*. Department of Earth Systems Analysis, University of Twente, Faculty of Geo-Information Science & Earth, Observation (ITC), Enschede (NL)

[12] Sarpong, S. A (2013). Modeling and Forecasting Maternal Mortality; an Application of ARIMA Models. *International Journal of Applied Science and Technology Vol. 3, No. 1; 19-28.*

[13] Sudeshna, G. (2017). Forecasting Cotton Exports in India using the ARIMA model. *Amity Journal of Economics, Vol. 2, Issue 2, 36-52*

[14] Wei, W. W. S. (2006). *Time Series Analysis Univariate and Multivariate Methods*. Addison Wesley Publishing Company, Inc. Canada

[15] Zul Amry and Siregar, B.H. (2019). ARIMA Model Selection for Composite Stock Price Index in Indonesia Stock Exchange, *International Journal of Accounting and Finance Studies, Vol. 2, No. 1, 31-38.* https://doi.org/10.22158/ijafs.v2n1p31