

# Forecasting of Revenue, Number of Plane Movements and Number of Passenger Movements at Sultan Iskandar Muda International Airport Using the VARIMA Method

Asep Rusyana<sup>1\*</sup>, Lia Rahmati<sup>2</sup>, and Nurhasanah<sup>3</sup>  
asep.rusyana@unsyiah.ac.id<sup>1</sup>, lia.rahmati@students.stat.unsyiah.ac.id<sup>2</sup>, nurhasanah.math@gmail.com<sup>3</sup>

Department of Statistics, Faculty of Mathematics and Natural Sciences, Universitas Syiah Kuala,  
Jalan Syech Abdurrauf No.3, Kopelma Darussalam, Banda Aceh, Aceh, Indonesia 23111<sup>1,2,3</sup>

**Abstract.** Forecasting has become a necessity in various fields. One of the companies that do the forecasting is PT. Angkasa Pura II - Sultan Iskandar Muda (SIM) International Airport. The amount of income, the number of aircraft movements, and the number of passenger movements are some of the interesting things to study from an airport. This kind of problem can be solved by Vector Autoregressive Integrated Moving Average (VARIMA) method. The variables involved in this data analysis can be modeled and forecasted. The analysis shows that the three variables studied tend to have a flat trend, where the amount of income, the number of aircraft movements, and the number of passenger movements has increased over time. The best model obtained is in the form of VARIMA (1,1,0) model. Percentage of forecast errors for income variables, aircraft movements, and passenger movements amounted to 14.028%, 11.003% and 13.330%, respectively.

**Keywords:** Aircraft Movement, Forecasting, Passenger Movements, Revenue, VARIMA.

## 1 Introduction

Forecast data has become a necessity for various interests. The data is usually used as a basis for designing a policy and decision to obtain more measurable results. Revenue is one of the important variables for forecasting, it also happens at PT. Angkasa Pura II, Sultan Iskandar Muda International Airport. The revenue of an airport can be obtained from several sources, such as the number of aircraft movements and the number of passenger movements.

There are several methods of forecasting. They are used according to characteristic of a data and ways. For illustration, SARIMA methods is an ARIMA developing which suitable for seasonal data [1]. Another example, decomposition method which has assumption that data consist of seasonal, cyclical, trend and error, so it forecasts future data with these components [2]. In the study, There are income variable, aircraft movements and passenger movements in the airport will be forecasted simultaneous, therefore this problem can be solved using the Vector Autoregressive Integrated Moving Average (VARIMA) method. The method can take into account data or data errors from the variables themselves and other variables at once. Therefore, the revenue

contributing variables such as the number of aircraft movements and the number of passenger movements can also be made modeling and forecasting simultaneously for each of them.

Based on the description above, there are two purposes of this study. The first purpose is to get a forecasting model for income, the number of aircraft movements, and the number of passenger movements. The second is to predict the variables during three years in the future. This research is expected to be a reference in the design of policies in SIM airport, and can provide information about VARIMA modeling.

## 2 Materials

### 2.1 Time series data

Time series data is a data set of an observation  $Y$  over time. Data relating to this time period has its own analysis method, which is time series analysis. Time series data has characteristics in the form of dependence between data over time. This characteristic is usually owned by single variable data. But a time series data variable sometimes also has additional influences from other variables. This kind of problem can be overcome by analyzing the time series of multiple variables.

### 2.2 Stationarity

Stationarity is one of the requirements of time series data analysis. Stationarity will occur if there is no growth or decrease in data [3]. Data is expected to be horizontal on the time axis. Where data fluctuations are constantly around the average value, not time dependent, and have constant variance. Stationarity in time series data is two kinds, namely stationary in variance and stationary in mean. Stationarity in variance can be predicted through the BoxCox lambda value as in equation (1) [4]. Data is said to be stationary in variance if  $\lambda = 1$ . But if  $\lambda \neq 1$  then there is a problem of data stationary in variance. This problem can be overcome through power transformations.

$$L(\lambda) = -\frac{1}{2} n \log\{S(\lambda; z)\} / n(1)$$

where  $S(\lambda; z)$  is the sum of squares of errors from  $z^\lambda$  and  $z$  is the normal standard form of variable  $Y$ .

The stationarity in mean can be tested using the Augmented Dickey Fuller (ADF) test. The ADF statistical values can be calculated by equation (2) [5] and the hypothesis for the test is as follows:

$H_0: \rho = 1$  (data contain unit root or data is not stationary)

$H_1: \rho \neq 1$  (data unit does not contain unit roots or data is stationary)

If the results are not stationary, then it can be overcome by differencing with formulas (3) [3].

$$t_{hit} = \frac{\hat{\rho}-1}{\sigma_{\hat{\rho}}} \quad (2)$$

$$y'_t = y_t - y_{t-1} \quad (3)$$

### 2.3 Vector autoregressive integrated moving average (VARIMA)

This VARIMA model can be applied to identify relationship between the observations and their errors in the previous time period. This method is also able to make model of data with mean stationary problems. The following is a mathematical formula for the VARIMA model [6].

$$\Phi_p(B)(I - B)^d Y_t = \Phi_0 + \Theta_q(B)a_t \quad (4)$$

or

$$Y_t = (I + \Phi_1)Y_{t-1} + (\Phi_2 - \Phi_1)Y_{t-2} + \dots + (\Phi_p - \Phi_{p-1})Y_{t-p} - \Phi_p Y_{t-p-1} + a_t - \Theta_1 a_{t-1} - \dots - \Theta_q a_{t-q}$$

### 2.4 Diagnostic model parameters

Diagnostic model parameters include two stages, namely the suitability of the model and the test of significance parameters. Suitability of the model can be proven through white noise testing and error normality testing. White noise statistics can be calculated by equation (5) [7]. The hypothesis formulation is as follows:

H<sub>0</sub>:  $\Gamma_i = 0$  (galat *white noise*)

H<sub>1</sub>:  $\Gamma_i \neq 0$  (galat tidak *white noise*).

$$\hat{Q}_k = T^2 \sum_{i=1}^k \frac{1}{T-i} \text{tr}[\hat{\Gamma}_{\hat{a}}(i)\hat{\Gamma}_{\hat{a}}^{-1}(0)\hat{\Gamma}'_{\hat{a}}(i)\hat{\Gamma}_{\hat{a}}^{-1}(0)] \quad (5)$$

Where  $\hat{\Gamma}_{\hat{a}}(i) = \frac{1}{T} \sum_{t=1}^{T-1} \hat{a}_t \hat{a}'_{t+i}$  ;  $i = 0, 1, \dots, k$  and  $k$  is the number of covariance matrix lags from errors, and  $T$  is the number of observations ( $t = 1, 2, \dots, T$ ). The  $\hat{Q}_k$  statistic corresponds to the distribution  $\chi^2$  with degree of freedom  $(k-p-q)$ .

Normality error was tested by the Mardia test as in equation (6) [8]. The hypothesis for this test is as follows:

H<sub>0</sub>:  $a_t = 0$  (normal double variable error spread)

H<sub>1</sub>:  $a_t \neq 0$  (error does not spread the normal double variable).

$$\hat{\gamma}_{1,p} = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n m_{ij}^3 \text{ and } \hat{\gamma}_{2,p} = \frac{1}{n} \sum_{i=1}^n m_{ii}^2 \quad (6)$$

where  $m_{ij} = (y_i - \bar{y})' S^{-1} (y_j - \bar{y})$  and  $p$ : number of variables.

Testing the significance of model parameters can be measured through the t test (7). This test is done by knowing the significance of the parameters in the model. The following hypothesis for the test.

H<sub>0</sub>:  $\theta_{i,j} = 0$  (parameter not significant)

H<sub>1</sub>:  $\theta_{i,j} \neq 0$  (significant parameter).

$$t_{hit} = \frac{\hat{\theta}_{i,j}}{SE(\hat{\theta}_{i,j})} \quad (7)$$

where  $\hat{\theta}_{i,j}$  is an estimate for the parameter  $\theta_{i,j}$ .

## 2.5 Prediction accuracy measures

The level of accuracy of each forecast can be measured by measuring the accuracy of forecasting. Measuring instruments that can be used for measuring the accuracy of forecasting are as follows [3].

a. *Mean Absolute Error* (MAE)

$$MAE = \frac{1}{n} \sum_{t=1}^n |Y_t - \hat{Y}_t| \quad (8)$$

b. *Mean Absolute Percentage Error* (MAPE)

$$MAPE = \frac{100}{n} \sum_{t=1}^n \left| \frac{Y_t - \hat{Y}_t}{Y_t} \right| \quad (9)$$

c. *Mean Square Error* (MSE)

$$MSE = \frac{1}{n} \sum_{t=1}^n (Y_t - \hat{Y}_t)^2 \quad (10)$$

## 2.6 Airport revenue contributors

The income data used in this study include total income from the income of aeronatics services and income from non-aeronautical services. However, the Airport revenue forecasting modeling only involves contributing factors to revenue from aeronatics services only, in the form of the number of aircraft movements and the number of passenger movements. The explanation for the two variables is as follows.

a. Number of aircraft movements

The aircraft movement variables used in this study include all movements of aircraft that land on SIM airports, both the movements of civilian and military aircraft and both domestic and foreign aircraft movements. The data is in the form of the total movement of aircraft landing, both who arrive and depart for each month.

b. Number of passenger movements

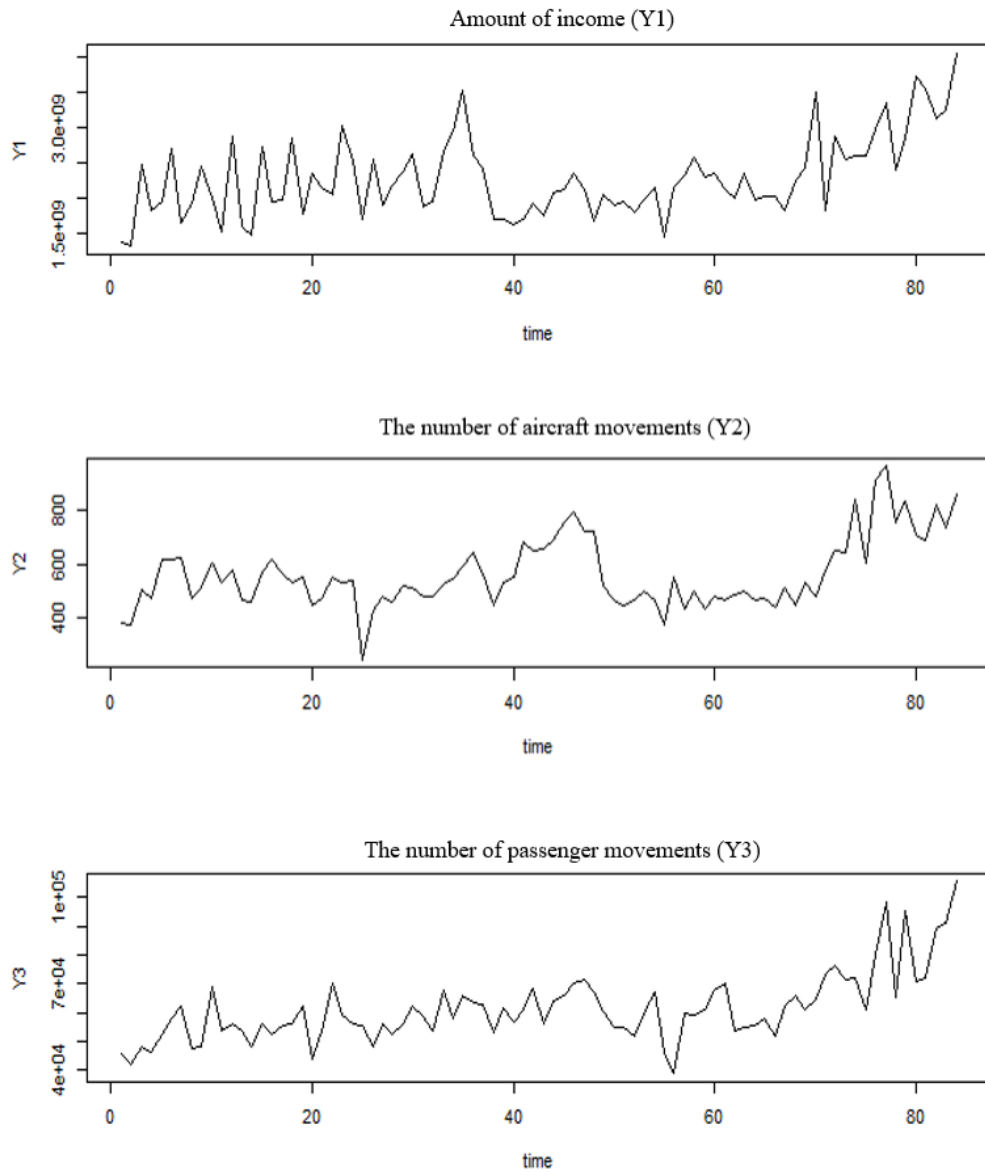
The passenger movement variables analyzed in this study are the total movements of airplane passengers at the SIM Airport. The number of passenger movements includes the number of movements of passengers who arrive and depart for each month. The number of passenger movements also includes domestic and international passengers.

### 3 Methods

The data used in this study is monthly from January 2010 to December 2016. The data is obtained from PT. Angkasa Pura II Branch Office of Sultan Iskandar Muda International Airport (SIM), Aceh. There are three variables analyzed, namely the amount of income, the number of aircraft movements, and the number of passenger movements. The data used in this study is divided into two parts, namely training data and testing data. Training data is used to build forecasting model. The training data in this study are in 3.5 years of data, from 2010 to mid-2013 or 42 observations. While testing data is used to measure the ability of the model in forecasting. The testing data also consists of 42 observations which are data from July 2013 to December 2016. The method used in this study is the Vector Autoregressive Integrated Moving Average (VARIMA).

### 4 Results and Discussion

**Figure 1** shows the plot of the time series data for the three variables used. The three data patterns in **Figure 1** have almost the same pattern. The three variables tend to have a trend pattern, the three variables have a sharp trend in the final period or around 2016. The similarity of data patterns make the VARIMA method is applied for data analysis of the three variables. while the VARIMA modeling stages are as follows.



**Fig. 1.** Time series data plot.

a. Stationary testing

The stationarity tested is two, namely stationary in variance and stationary in mean. Stationarity testing in variance is done using BoxCox Lambda. Table 1 shows lambda values obtained.

**Table 1.** BoxCox lambda values.

Variable	Lambda Value
Y <sub>1</sub>	-0.999
Y <sub>2</sub>	1.999
Y <sub>3</sub>	-0.999

**Table 1** shows that there is no variable that has a lambda = 1 or that has been stationary, so that all variables must be transformed. The transformation carried out in the form of power transformation. The following is a transformation and the lambda value obtained.

**Table 2.** Types of transformation and lambda values after transformation.

Variable	Transformation Type	Lambda Value
Y <sub>1</sub>	$Y_1^* = Y_1^{-0.159} \times 1000$	1.015
Y <sub>2</sub>	$Y_2^* = Y_2^{-0.178} \times 1000$	1.009
Y <sub>3</sub>	$Y_3^* = Y_3^{2.61} \times 10.000.000$	1.002

After the transformation is done, the data becomes stationary. This is known from the lambda value in Table 2, where all lambda values for each variable are close to 1. The next step is to test stationarity in the mean.

Stationarity is tested on average using the Augmented Dickey Fuller Test (ADF). The results obtained are as follows.

**Table 3.** Statistical value of Augmented Dickey Fuller Test.

Variable	T	P-value
Y <sub>1</sub>	-3.877	0.100
Y <sub>2</sub>	-2.191	0.233
Y <sub>3</sub>	-3.592	0.010

Table 3 shows that there is one variable that has been stationary, namely the variable Y<sub>3</sub>. This can be seen from the p-value held smaller than alpha (0.01 < 0.05). While variables Y<sub>1</sub> and Y<sub>2</sub> are not stationary, where p-value > alpha (0.1 > 0.05 and 0.233 > 0.05). The variable that is not stationary can be modified by differencing. The results obtained are in the form of stationary data with only one differencing time.

b. Estimation and testing of parameters and model diagnostics

After parameter estimation is done for several models, only the VARIMA (1,1,0) model can be estimated by its parameters. The next step is to do a diagnostic model that includes white noise test and error normality test. The results obtained for both tests are as follows.

**Table 4.** Results of error normality and white noise test.

	Mardia Test		White Noise Test	
	Statistic	<i>P-value</i>	$Q_m$	<i>P-value</i>
Mardia Skewness	13.264	0.210	7.930	0.541
Mardia Kurtosis	-0.020	0.980		

Table 4 shows that the VARIMA (1,1,0) model meets the requirements for error error and white noise. This was decided based on  $p\text{-value} > \alpha (0.05)$  for the two tests. Therefore, the VARIMA (1,1,0) model is suitable for forecasting. Furthermore, the model is tested by parameters. The results obtained are as follows.

**Table 5.** Estimated values and parameter testing.

Parameter	Estimated Value	Std error	T-statistic	<i>p-value</i>
$Y_1^*1$	-0.487	0.146	-3.325	0.0009
$Y_2^*1$	-0.317	0.583	-0.543	0.587
$Y_3^*1$	-0.105	0.073	-1.434	0.151
$Y_1^*2$	-0.059	0.049	-1.200	0.230
$Y_2^*2$	-0.174	0.197	-0.880	0.379
$Y_3^*2$	0.023	0.025	0.933	0.351
$Y_1^*3$	0.504	0.320	1.575	0.115
$Y_2^*3$	-2.373	1.275	-1.862	0.063
$Y_3^*3$	-0.621	0.160	-3.868	0.0001

A parameter is said to be significant if  $p\text{-value} < \alpha (0.05)$ . Table 5 shows that there are two parameters that are significant, namely  $Y_1^*1$  and  $Y_3^*3$ . This shows that only these two parameters affect the model. The VARIMA (1,1,0) model can be written as follows.

$$\begin{bmatrix} \hat{Y}_{1,t} \\ \hat{Y}_{2,t} \\ \hat{Y}_{3,t} \end{bmatrix} = \begin{bmatrix} 0.513 & -0.035 & 0.309 \\ -0.050 & 0.826 & -0.162 \\ -0.189 & 0.498 & 0.379 \end{bmatrix} \begin{bmatrix} Y_{1,t-1} \\ Y_{2,t-1} \\ Y_{3,t-1} \end{bmatrix} - \begin{bmatrix} -0.484 & -0.035 & 0.309 \\ -0.050 & -0.171 & -0.162 \\ -0.189 & 0.498 & -0.465 \end{bmatrix} \begin{bmatrix} Y_{1,t-2} \\ Y_{2,t-2} \\ Y_{3,t-2} \end{bmatrix}$$

The model is broken down into three equations for each variable. The three equations are as follows:



$$\begin{aligned}\hat{Y}_{1,t} &= 0.513Y_{1,t-1} - 0.035Y_{2,t-1} + 0.309Y_{3,t-1} + 0.484Y_{1,t-2} + 0.035Y_{2,t-2} - 0.309Y_{3,t-2} \\ \hat{Y}_{2,t} &= -0.05Y_{1,t-1} + 0.826Y_{2,t-1} - 0.162Y_{3,t-1} + 0.05Y_{1,t-2} + 0.171Y_{2,t-2} + 0.162Y_{3,t-2} \\ \hat{Y}_{3,t} &= -0.189Y_{1,t-1} + 0.498Y_{2,t-1} + 0.379Y_{3,t-1} + 0.189Y_{1,t-2} - 0.498Y_{2,t-2} + 0.465Y_{3,t-2}.\end{aligned}$$

The three equations can be used for forecasting each variable.

#### c. Measurement of forecast accuracy

There are three measuring instruments used in measuring the accuracy of predictions, namely Mean Square Error (MSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE). The results for these measurements are as follows.

**Table 6.** Measurement of forecast accuracy.

Variable	Measurement Tool		
	MSE	MAE	MAPE
Y <sub>1</sub>	$2.037 \times 10^{17}$	345,628,063.9	14.028%
Y <sub>2</sub>	8,320.32	69.06	11.003%
Y <sub>3</sub>	$1.2 \times 10^8$	8,896.41	13.330%

Table 6 shows the size of forecast accuracy for each variable. The average of the forecasting error variable Y<sub>1</sub> (amount of income) is  $2.037 \times 10^{17}$ . While the average absolute error is 345,628,063.9 and the mean absolute percentage of errors is 14.028%. The variable Y<sub>2</sub> or the number of aircraft movements has MSE, MAE, and MAPE values of 8,320.32; 69.06; and 11.003%, respectively. While the error rate for variable Y<sub>3</sub> or the number of passenger movements is  $1,2 \times 10^8$  for MSE, 8,896.41 for MAE, and 13.330% for MAPE. Based on the description above, it can be seen that the largest percentage of forecast errors occur in the income variable forecast which is equal to 14.028%. Whereas the forecast results for aircraft movement variables have a smaller error rate than the other two variables.

#### d. Data forecasting

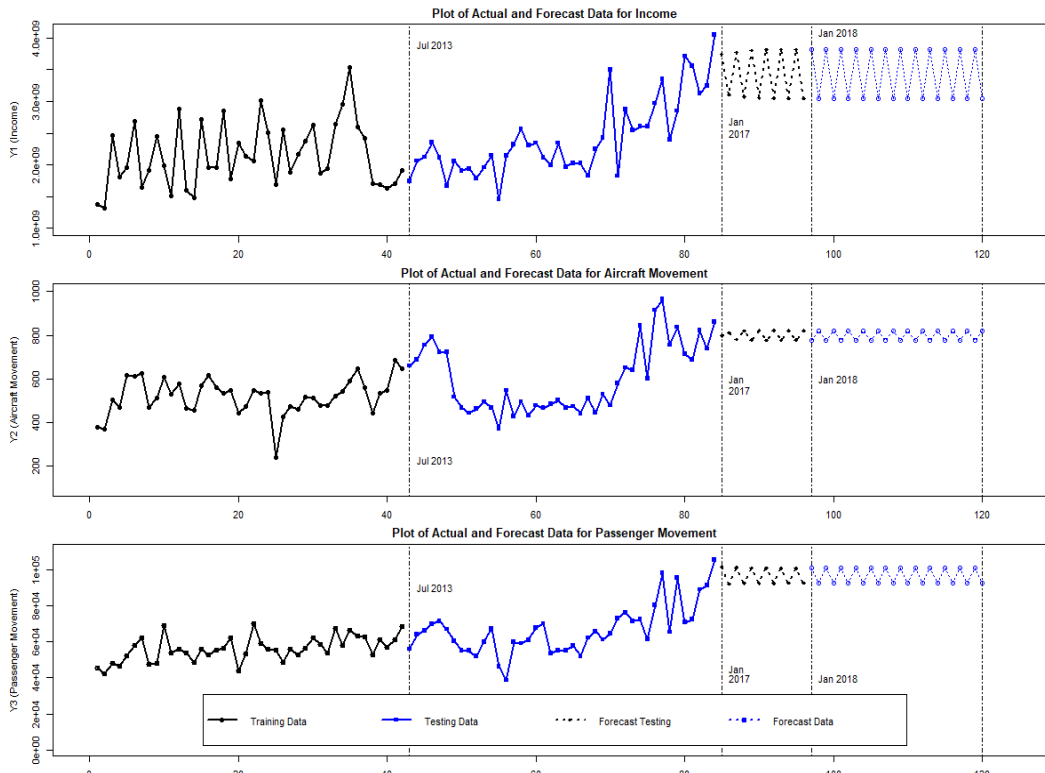
The forecast results obtained from the best models are as follows:

**Table 7.** Forecasting data.

Period	Income	The number of aircraft movement	The number of passenger movement
Jan-17	3,745,354,668	798	101,686
Feb-17	3,102,351,412	812	92,085
Mar-17	3,772,836,982	781	101,242
Apr-17	3,069,992,587	820	92,357
May-17	3,805,132,198	778	101,016

Period	Income	The number of aircraft movement	The number of passenger movement
Jun-17	3,053,997,182	821	92,497
Jul-17	3,816,620,810	777	100,920
Aug-17	3,049,518,247	821	92,546
Sep-17	3,819,286,209	778	100,891
Oct-17	3,048,670,742	821	92,559
Nov-17	3,819,676,918	778	100,885
Dec-17	3,048,591,508	821	92,561
Jan-18	3,819,680,524	778	100,884
Feb-18	3,048,607,380	821	92,561
Mar-18	3,819,660,779	778	100,884
Apr-18	3,048,617,631	821	92,561
May-18	3,819,653,160	778	100,884
Jun-18	3,048,620,626	821	92,561
Jul-18	3,819,651,366	778	100,884
Aug-18	3,048,621,196	821	92,561
Sep-18	3,819,651,105	778	100,884
Oct-18	3,048,621,247	821	92,561
Nov-18	3,819,651,105	778	100,884
Dec-18	3,048,621,235	821	92,561
Jan-19	3,819,651,119	778	100,884
Feb-19	3,048,621,228	821	92,561
Mar-19	3,819,651,124	778	100,884
Apr-19	3,048,621,226	821	92,561
May-19	3,819,651,126	778	100,884
Jun-19	3,048,621,226	821	92,561
Jul-19	3,819,651,126	778	100,884
Aug-19	3,048,621,226	821	92,561
Sep-19	3,819,651,126	778	100,884
Oct-19	3,048,621,226	821	92,561
Nov-19	3,819,651,126	778	100,884
Dec-19	3,048,621,226	821	92,561

The following is a visualization of forecast data Table 7.



**Fig. 2.** Actual Data Testing Plots and Forecast Data.

In general, the forecasting model obtained follows the movement of the actual data, see **Figure 2**. However, some forecast data points appear far from the actual data. This may be influenced by the randomness factor composition in the data. While forecast data for the next 12 periods form a fluctuating pattern. Forecast data has increased and decreased over time. The highest fluctuations occur in the variable  $Y_1$  forecast data or income, while the forecast data for  $Y_2$  variable or the number of passenger movements experiences the lowest fluctuations compared to the other three variables.

## 5 Conclusions and Recommendations

Based on the results and discussion above, there are two conclusions of the study:  
 (1) VARIMA (1,1,0) is the best model resulted in the study.

- (2) Data of forecasting for three years are resulted with VARIMA (1,1,0) model .it can be seen that the pattern of data from income variable, the number of aircraft movements, and the number of passenger movements tend to have a similar pattern. These three variables have patterns that tend to be flat trends.

This research results are expected to be a reference in the design of policies in SIM airport.

## References

- [1] Rusyana, A., Nurhasanah, Marzuki and Flancia, M.: SARIMA model for forecasting foreign tourists at the Kuala Namu International Airport In: The 12th International Conference Mathematics, Statistics, and their Applications (ICMSA). Banda Aceh, Indonesia: IEEE p. 153–158 <https://ieeexplore.ieee.org/document/7954329/> (2016)
- [2] Rusyana, A., Nurhasanah and Astuti, S.: Forecasting the number of visitors of Aceh state museum using decomposition method In: Journal of Physics: Conference Series (2018)
- [3] Makridakis, S., Wheelwright, S. C. and McGEE, V. E.: Forecasting: Methods and Applications 2nd ed. New York: Wiley (1983)
- [4] Box, G.E.P. and Cox, D.R: An Analysis of Transformation J. R. Stat. Soc. Ser. B. **26** 211–252 (1964)
- [5] Hamilton, J.D.: Time Series Analysis. New Jersey: Princeton University Press (1994)
- [6] Wei, W.W.S.: Time Series Analysis Univariate and Multivariate Methods 2nd ed. Boston: Pearson Education, Inc.(2006)
- [7] Box, G.E.P., Jenkins, G.M. and Reinsel, G.C.: Time Series Analysis: Forecasting and Control 4th ed. New Jersey: A John Wiley & Sons, Inc. (2008)
- [8] Korkmaz, S., Goksuluk, D. and Zararsiz, G.: MVN: An R Package for Assessing Multivariate Normality. Ankara: Hacettepe University (2016)