Forecasting and Effectiveness Analysis of Domestic Airplane Passengers in Yogyakarta Adisutjipto Airport with Autoregressive Integrated Moving Average with Exogeneous (ARIMAX) Model

Sarah Khairunnisa¹, Nusyrotus Sa'dah², Isnani¹, Rohmah Artika¹, Prihantini¹

¹Mathematics, ²Statistics, Faculty of Mathematics and Natural Science, Yogyakarta State University Jl. Colombo No.1, Karang Malang, Caturtunggal, Kec. Depok, Kabupaten Sleman, Daerah Istimewa Yogyakarta 55281

Abstract. Airplane is one of the public transportations options that many people choose when traveling long distance. Nowadays, it is notes that the number of passengers domestic flight has increased from the previous months. This increase, especially occurs on the holidays, such as year-end holidays, Eid, and others. The increase of airplane passengers is inversely proportional to the number of available airplane. Forecasting the number of airplane passengers is necessary to prepare additional facilities when there is increasing passengers. This research focused on forecasting domestic airplane passengers at Adisucipto Airport, Yogyakarta using ARIMAX method to forecast the number of domestic airplane passengers and the effectiveness of domestic passengers at the international airport. The purpose of this research is to determine the best ARIMAX model and forecast airplane passengers in Adisucipto airport. The results will show the effectiveness of ARIMAX model with the effect of calendar variance on domestic airplane passenger forecasting at international airport. Based on the result of AIC and RMSE values, it shows that the ARIMAX(1,0,1) model with calendar variation is better than ARIMA(1,0,1) in predicting the number of airplane passengers at Yogyakarta Adisutjipto airport.

Keywords: ARIMAX Model, Effectiveness Analysis, Forecasting

Abbreviations: ARIMAX (Autoregressive Integrated Moving Average with Exogeneous)

INTRODUCTION

The increasing number of population causes the increasing use of transportation services. One of the transportations that many people choose is airplane. That is because, if you are travelling by airplane, the time needed to get to the destination becomes faster. Airplane is also one of the transportation that many people choose when traveling long distances. The increase in airplane passengers mainly occurs during holidays, such as yearend holidays, Eid and others. Badan Pusat Statistika (BPS) recorded the number of airplane passengers for domestic flights in October 2018 increased 6,85% to 8,11 million people and also grew 7,85% compared to October of the previous year. Likewise, cumulatively in January-October period this year grew 6,98% to 78,63 million people compared to the same period in 2017 (Databoks, 2018). The increasing number of airplane passengers must be followed by the availability of the airplant fleet, so there is no long queues.

Yogyakarta Adisutjipto Airport is one of the international airports that has an important role in the development of transportation and tourism in Indonesia. The number of passengers at the Adisutjipto Airport has increased every year, especially during the holidays and the end of the year. PT Angkasa Pura I (Persero) Yogyakarta Adisutjipto International Airport recorded passenger growth during 2018 is 8.417.089, up 7,65% compared to 2017 as 7.818.871 passengers (Adisutjipto International Airport, 2019). The effect of the Eid day also had an effect on increasing the number of airplane Yogyakarta passengers at Adisutjipto Airport. Passengers number data is time series data that is collected, recorded, and observed every year to determine the increase number of passengers at Yogyakarta Adisutjipto Airport. Therefore, it is necessary to model and forecast the number of airplane passengers at Yogyakarta Adisutjipto Airport. This study aims to determine the forecasting of the number of domestic airplane passengers at PT. Angkasa Pura I (Persero) Yogyakarta Adisutjipto International Airport using the Autoregressive Integrated Moving Average with Exogeneous (ARIMAX) method with calendar variations. The data used are secondary data obtained from the Transportation Department of Special Region of Yogyakarta in January 2013 to August 2018.

Research related to forecasting the number of passengers was also carried out by Juniar Iqbalullah and Wiwiek Setya Winahju in 2014 with forecasting the number of airplane passengers at the arrival gate of the Lombok International Airport by the Arima Box-Jenkins, Arimax, and time series regression methods. In addition, Yoan Sofyana Noer, Gumgum Darmawan, Resa Septiani Pontoh in 2016 forecasted the number of passengers in Fajar Utama Train of Yogyakarta - Pasar Senen using variations in the ARIMAX calendar (Case Study at PT. Kereta Api Indonesia). Based on previous studies, the ARIMAX method with calendar variations is suitable to forecast the number of airplane passengers at Yogyakarta Adisutjipto Airport with time series data patterns that are affected by calendar variations.

METHODS

The data used in this study are secondary data obtained from the Transportation Department of Special Region of Yogyakarta in January 2013 to August 2018. The Transportation Department of Special Region of Yogyakarta located at Babarsari Street No.30, Janti, Caturtunggal, Depok, Sleman Regency, Special Region of Yogyakarta 55281.

Auto Regressive Integrated Moving Average (ARIMA)

ARIMA is one of the forecasting methods that have been introduced by G.E.P. Box and G. Jenkins. ARIMA models are a class of linear models that is capable of representing stationary as well as non-stationary time series (E. Hanke & W. Wichern, 2006). There are several models that have been produced using the method Box-Jenkins is a moving average (MA), autoregressive (AR) model, one class model useful for time series which is a combination of MA and AR processes called ARMA. These models are linear and stationary models of the Box-Jenkins method. While the model for non-standard data is the ARIMA model. There are four groups of ARIMA classification, are: Autoregressive (AR), Moving Average (MA) and mixed model Autoregressive Integrated Moving Average (ARIMA) which has characteristics from the first two models as well as the Seasonal Autoregressive Integrated Moving Average (SARIMA) which is a derivative of ARIMA to get seasonal data predictions.

a. Autoregressive (AR)

The general form of the autoregressive model with the order p(AR(P)) or the ARIMA model (p, 0.0) say as follows :

 $x_t = \mu' + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + e_t[0]$ Where:

 $\mu = a \text{ constant}$

b. Moving average Model (MA)

The order model q (*MA* (q)) or ARIMA (0.0, q) as follows :

 $X_t = \mu' + e_t - \phi_1 e_{t-2} - \dots - \phi_q e_{t-k}$

Where:

 $\mu = a \text{ constant}$

 $\phi_1 - \phi_q$ = parameter moving average

- e_{t-k} = error value-t k
- c. Mixed Model

i. ARMA process

The general model for a mixture of pure AR (1) and MA (1) processes, for example ARIMA (1,0,1) as follows :

$$\begin{aligned} X_t &= \mu' + \phi_1 X_{t-1} + e_t - \phi_1 e_{t-1} \\ atau \\ (1 - \phi_1 B) X_t &= \mu' + (1 - \phi_1) e_t \end{aligned}$$

ii. ARIMA process

If non-stationarity is added to the ARMA process mixture, then the general ARIMA model (p, d, q) is fulfilled. The equation for the simple case of ARIMA (1,1,1) is as follows:

$$(1 - \phi_1 B)X_t = \mu' + (1 - \phi_1)e_t$$

d. Autocorelation Function (ACF)

ACF is the correlation between data in time period t with the previous time period t - 1. The mean and range of periodic data may not be useful if the series is not stationary, but the maximum and minimum values can be used for plotting purposes. How the key statistics in the periodical analysis are autocorrelation coefficients. The ACF equation can be seen in the following formula:

$$r_k = \frac{\sum_{t=b}^{n} (x_t - \bar{x}) (x_{t+k} - \bar{x})}{\sum_{t=b}^{n} (x_t - \bar{x})^2}$$

e. Partial Autocorelation Function (PACF)

Partial autocorrelation is used to measure the level of intelligence between x_t and x_{t-k} , if the effects of lag time are considered separately. The only objective in periodic series analysis is to help determine the right ARIMA model. Sample value the ordered PACF can be seen in the formula:

$$\begin{aligned} rkk &= r_{1} & \text{if } k = 1 \\ rkk &= \frac{rk - \sum_{j=1}^{k-1} r_{k-1,j} r_{k-j}}{k-1} & \text{if } k = 2,3 \\ rkk &= 1 - \sum_{j=1}^{k-1} r_{k-1,jr-j} & \end{aligned}$$

Auto Regressive Integrated Moving Average with Exogeneous (ARIMAX)

ARIMAX model is a modification of seasonal ARIMA. This modification is done by add predictor variables. Variations calender effects is one of the variables that are often used in the modeling. (wiwik, Vinarti, & Kurniati, 2015). There are two kind of additional variables, I.e. the dummy variable(s) for calendar variation effects only, and dummy variable(s) for calendar variation effects and deterministic trend. The first model is known as ARIMAX with stichastic trend by implementing a difference non seasonal or seasonal and the second with deterministic trend (without differencing order). (Lee, Suhartono, & Hamzah, 2010) ARIMAX model with stochastic trend is

$$y_{t} = \beta_{1}V_{1,t} + \beta_{2}V_{2,t} + \dots + \beta_{p}V_{p,t} + \frac{\theta_{q}(B)\theta_{Q}(B^{S})}{\phi_{p}(B)\phi_{p}(B^{S})(1-B)^{d}(1-B^{S})^{D}} \varepsilon_{t}$$

ARIMAX model with deterministic trend is

$$y_t = \gamma t + \beta_1 V_{1,t} + \beta_2 V_{2,t} + \dots + \beta_p V_{p,t} + \frac{\theta_q(B) \theta_Q(B^S)}{\emptyset_p(B) \Phi_p(B^S)^D} \varepsilon_t$$

The proposed ARIMAX model building procedure, in the presence of calendar variation effects in the time series data is described as follows:

- Step 1: determination of dummy variable for calender variation priode
- Step 2: remove the calendar variation effect from the response by fitting Eq. (2) for model with stochastic trend, or fitting Eq. (1) and (2) simultaneously for model with deterministic trend, in order to obtain the error, w_t .
- Step 3: model w_t using ARIMA model (see Boxjenkins procedure)
- Step 4: the order of ARIMA model found from step 3 is used to model the real data and dummy variables of calendar variation effect as input variables simultaneously as Eq. (5) and (6) for model with stochastic and determinisic trends respectively.
- Step 5: test parameter significance and perform diagnostic checks until the process is stationary and ε_t appears as white noise processes.

RESULTS AND DISCUSSION

The number of airplane passengers at Yogyakarta Adisutjipto Airport in January 2013 – August 2018 showed an upward trend in the month of Eid and the end of the year. Thus, the data can be built using ARIMAX models with dummy calendar variations when Eid (E_t) , the month before Eid (E_{t-1}) , the month after Eid (E_{t+1}) , the end of the year (T_t) , and the beginning of the year (A_t) .

Forecasting Model The Number of Passengers Airplane with ARIMA

In determining the forecast model the number of airplane passengers using the ARIMA Box-Jenkins method, can be seen from the time series plot of the number of aircraft passengers data to find out the pattern and characteristics of the data. The graph of the number of airplane passengers at Yogyakarta Aditsutjipto Airport in January 2013 – August 2018 can be seen in the Time Series Plot Figure 1.



Figure 1. Plot Data of Passengers Domestic Month Period January 2013 to August 2018.

Based on the plot in Figure 1 shows the number of airplane passengers in January 2013 is 446.817 passengers In July 2018 was the period with the highest number of passengers of 677.390 passengers and then decreased in August 2018 with a total of 58.283 passengers. From Figure 1, it can be seen that the data shows a trend or changing pattern. Therefore, it cannot be said that the data is stationary. Stationary data with respect to variants can be seen using Box-Cox Plot in figure 2. The rounded value is 1,00, it is means that data of the number of airplane passenger at Yogyakarta Adisutjipto Airport in January 2013 – August 2018 is stationary with respect to variants.



Figure 2. Box-Cox Plot Data of Passengers.

Stationary data with respect to means can be seen in the graph autocorrelation function (ACF) and partial autocorrelation function (PACF) as shown in Figure 3 and 4.



Figure 3. Autocorelation Funtion Data of Passengers Domestic



Figure 4. Partial Autocorelation Data of Passengers Domestic.

Based on Figure 3 and Figure 4 can be seen that the data has been stationary with respect to means because the only significant autocorrelation is in lag 1 and significant partial autocorrelation is also in lag 1. Autocorrelation was interrupted in lag 1, this shows that the MA (1). At the same time, partial autocorrelation appears to be interrupted after lag 1, this shows that the AR (1). ARIMA models that can be formed based on ACF and PACF are: ARIMA(1, 0, 0), ARIMA(0, 0, 1), and ARIMA(1, 0, 1). The following are the results of the ARIMA model calculation:

Tabel 1. The results of the ARIMA model calculation.

Model		Sig	AIC
ARIMA(1, 0, 1)	AR1 = 0.7901	0.000	- 1693.39
	MA1 = -0.4037	0.000	
ARIMA(1,0,0)	AR1 = 0,4602	0.000	1693.66
ARIMA(0,0,1)	AR1 = 0,3513	0.000	1697.99

The sig value for the three models is 0.000, it is means that the three models are significant. However, the smallest AIC value is in the ARIMA model (1,0,1), so it can be said that the model is the best ARIMA model used to model data on the number of passengers of domestic airplane at Yogyakarta Adisutjipto Airport.

Forecasting Model The Number of Passengers Airplane with ARIMAX

After obtaining the Arima model, then, determining the ARIMAX model uses calendar variations to predict the number of domestic aircraft passengers at Yogyakarta's Adisutipto Airport. The dummy variable used is Eid (E_t) , the month before Eid (E_{t-1}) , the month after Eid (E_{t+1}) , the end of the year (T_t) , and the beginning of the year (A_t) . ARIMAX models that can be formed based on ACF and PACF are: ARIMAX (1, 0, 0), ARIMAX(0, 0, 1), and ARIMAX(1, 0, 1). The following are the results of the ARIMAX model calculation:

Tabel 2. The results of the ARIMAX model calculation.

	Model	Sig	AIC
(1, 0, 1) –	AR1 = 0.9765	0.000	1688
	MA1 = -0.7838	0.000	
(1,0,0)	AR1 = 0.4225	0.0023	1693
(0,0,1)	MA1 = 0.323	0.0095	1696

Based on the result of the ARIMAX model calculation, the sig value for the three models are less then 0.05, it is means that the three models are significant. However, the smallest AIC value is in the ARIMAX model (1,0,1), so it can be said that the model is the best ARIMAX model used to model data on the number of passengers of domestic airplane at Yogyakarta Adisutjipto Airport.

Comparison of the Best Forecasting Model Number of Passengers Airplane between ARIMA and ARIMAX

After getting the best model for forecasting the number of airplane passengers with ARIMA Box-Jenkins and ARIMAX, then the best model is then compared to predict the number of airplane passengers at Yogyakarta Adisutjipto Airport by using the Root Mean Square Error (RMSE) and Akaike's Information Criterion (AIC) criteria as benchmarks. The smaller the AIC and RMSE values the better the model.

Tabel 3 The AIC and RMSE values.

Model	AIC	RMSE
ARIMA(1,0,1)	1693.39	55171
ARIMAX(1,0,1)	1688	52440.61

Based on the table above, it can be seen that the AIC and RMSE values of the ARIMAX(1,0,1) model are smaller than the ARIMA(1,0,1) model. It shows that the ARIMAX(1,0,1) model with calendar variation is better than ARIMA(1,0,1) in predicting the number of airplane passengers at Yogyakarta Adisutjipto Airport.

Forecasting Number of Passenger Airplanes with ARIMAX(1,0,1)

It is known that the best model for predicting the number of airplane passengers at Yogyakarta Adisutjipto Airport is the ARIMAX model (1,0,1). The results of forecasting the number of airplane passengers in the period September 2018 - July 2019 using the ARIMAX model (1,0,1) are given as follows.

Tabel 4. The result of forecasting using ARIMAX(1,0,1).

Date	ARIMAX	95% Limits	
Date		-	
September 2018	579094	Lower	Upper
Oktober 2018	560614	481446	717119
November 2018	546012	477541	721691
Desember 2018	534475	473773	726125
Januari 2019	525359	470131	730435
Februari 2019	518156	466602	734631
Maret 2019	512464	463178	738723
April 2019	507967	459850	742719
Mei 2019	504414	456611	746627
Juni 2019	501607	453453	750453
Juli 2019	497635	450373	754203

Discussion

In this study, the exogeneous variable used is calendar variation as Eid, the month before Eid, the month after Eid, the end of the year, and the beginning of the year, which gives an influence on forecasting results. For further research, it is expected to be able to use other exogenous variables, which of course also have an effect on forecasting.

CONCLUSION

This research focused on forecasting domestic airplane passengers at Adisucipto Airport, Yogyakarta using ARIMAX method to forecast the number of domestic airplane passengers and the effectiveness of domestic passengers at the international airport. The purpose of this research is to determine the best ARIMAX model and forecast airplane passengers in Adisucipto airport. The results will show the effectiveness of ARIMAX model with the effect of calendar variance on domestic airplane passenger forecasting at international airport. Based on the result of AIC and RMSE values, it shows that the ARIMAX(1,0,1) model with calendar variation is better than ARIMA(1,0,1) in predicting the number of airplane passengers at Yogyakarta Adisutjipto airport. For further research, it is expected to be able to use other exogenous variables, which of course also have an effect on forecasting.

ACKNOWLEDGEMENTS

This study was supported by Transportation Department of Special Region of Yogyakarta who gave us the data of the number of airplane domestic passengers at Yogyakarta Adisutjipto Airport in January 2013 to August 2018 and gave us the opportunity to proceed the data.

REFERENCES

- (2018, Desember 20). Retrieved November 15, 2019, from Databoks: https://databoks.katadata.co.id/datapublish/2018/12/20/penum pang-pesawat-penerbangan-domestik-januari-oktober-2018tumbuh-7
- (2019, Januari 1). Retrieved November 18, 2019, from Adisutjipto International Airport: https://adisutjiptoairport.co.id/en/news/index/angkasa-pura-i-bandaraadisutjipto-catat-pertumbuhan-penumpang-hingga-7-65-ditahun-2018
- E. Hanke, J., & W. Wichern, D. 2006. *Bussines Forecasting*. New Jersey: Pearson Prentice Hall.
- Lee, M. H., Suhartono, & Hamzah, N. A. 2010. Calender variation model based on ARIMA for forecasting sales data with ramadhan effect. *Malaysia Institute of Statistics, Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA (UITM), Malaysia.*
- wiwik, A., Vinarti, R. A., & Kurniati, Y. D. 2015. performance Comparisons Between Arima and Arimax mothod in moslem kids clothes demand forecasting: case study. *Procedia Computer Science*, 630-637.

THIS PAGE INTENTIONALLY LEFT BLANK