

# Prediction of the Number of Passengers at Yogyakarta Airport

Imam Safawi Ahmad, Agus Suharsono and Elly Pusporini

**Abstract**—The development of air transportation services is growing up. Based on the report of Central Bureau of Statistics (BPS) in January-October 2015, the number of passengers reached 67.5 million. This number is increased by 12.8% from the previous year with only 58.9 million. Overall, there was increased in the number of passengers caused overload of capacity. One of international airports in Indonesia is Adisutjipto in Yogyakarta. The airport officer wants to develop in terms of airplane schedule management and the number of aircrafts used to calculate the number of passengers in the future. In this study, we aimed to forecast next period by using three methods, namely ARIMA, Exponential Smoothing and Combination of both univariate models. This research gives results that generally all route significantly increased every year with Denpasar, Jakarta, Pontianak and Singapore as exception. They were declined slightly in 2015. The number of passengers of departure and arrival routes are affected by seasonal impact. In addition, the model for departure and arrival data had similar models. Another result is combination method did not produce better results than the univariate method. The fit model for predicting data passenger is ARIMA seasonal methods.

**Index Terms**—ARIMA, exponential smoothing, forecasting.

## I. INTRODUCTION

AIR transportation becomes alternative mode covered long distance and shorter time consumption. The number of passengers is increasing across the time. It is also occurred at Adisutjipto airport Yogyakarta, which is one of the international airports in Indonesia. This causes the existing flight schedules at Adisutjipto Airport to be congested. It is known that capacity of the airport is under 2 million per year, but in fact in 2013 the passengers reached over 5.3 million and rose to 6.2 million in 2014 dramatically. The impact of overcapacity exceeded less services. Therefore, the Adisutjipto airport officer desperately needed a projection of the number of passengers in order to base line making strategic policy. To solve this overcapacity, government has a plan to relocate airport from the current place (Sleman) to Kulonprogo. One of regarding new strategic services is forecasting of passengers in the future in order to synchronize the capacity needed.

To predict the number of passengers at Adisutjipto Airport, we can use the ARIMA (Autoregressive Integrated Moving Average) method, Exponential Smoothing and the Combination of Variances-Covariance between the two methods. For the previous research, Bougas stated that the combination forecast method of individual models is always better than one

of the worst individual models [1]. This result is also supported by the third result of [2], which states that “The accuracy of the combination of various methods outperforms, on average, the specific methods being combined and does well in comparison with other methods.” This study aims to predict the aircraft passenger using secondary data that consists of departure and arrival data for both domestic and international cases.

## II. LITERATURE REVIEW

### A. Autoregressive Integrated Moving Average (ARIMA) Model

Autoregressive Integrated Moving Average (ARIMA) is one of the forecast methods for time series data. It is known that data which will be used for this method must be stationary w.r.t. variance and means. If the variance data is not stationary, then transformation of data must be done and if the means is the one which is not stationary then the differencing must be done [3]. ARIMA( $p,d,q$ ) models can be written as follow:

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

and for Seasonal ARIMA models can be written as follows

$$\phi_p(B^S)(1-B^S)^D Z_t = \Theta_Q(B^S)a_t.$$

Box and Jenkins proposed four steps of this method are identifications models, parameter estimate, diagnostic checking and the last is forecasting. In this research, the estimation method used is Maximum Likelihood (ML).

### B. Exponential Smoothing

Exponential Smoothing is a forecasting method that used constants that assign weight to current demand and previous forecast to arrive at new forecast [4]. Simple Exponential Smoothing was used if the time series data does not have trend or seasonal pattern with the forecast can be written as follow [5]

$$\hat{Z}_{t+1} = \hat{Z}_t + \alpha(Z_t - \hat{Z}_t),$$

where  $\alpha$  is a constant between 0 and 1. The other method is Holts linear method where the modification from Simple Exponential Smoothing which can be used to forecast data with trend included [5]. Forecasting with this method uses two smoothing constants  $\alpha$  (level) and  $\beta$  (trend) with the value between 0 and 1 and the forecast can be written as follow

- Level:  $\ell_t = \alpha Z_t + (1 - \alpha)(\ell_{t-1} + b_{t-1})$
- Trend:  $b_t = \beta^*(\ell_t - \ell_{t-1}) + (1 - \beta^*)b_{t-1}$
- Forecast:  $\hat{Z}_{t+h|t} = \ell_t + b_t h$

where  $\ell_t$  is the estimation of level of the series at  $t$  and  $b_t$  is the estimation from the slope of the series at  $t$ . The other

Manuscript received September 10, 2019; accepted November 20, 2019.

The authors are with the Department of Statistics, Institut Teknologi Sepuluh Nopember, Surabaya 60111, Indonesia. E-mail: safawi@statistika.its.ac.id, agus\_s@statistika.its.ac.id

Exponential Smoothing Method is Holt Winters Method which has two forecast models that is as follow

#### 1) Multiplicative Seasonality

- Level:  $\ell_t = \alpha \frac{Z_t}{s_{t-m}} + (1 - \alpha)(\ell_{t-1} + b_{t-1})$
- Trend:  $b_t = \beta^*(\ell_t - \ell_{t-1}) + (1 - \beta^*)b_{t-1}$
- Seasonal:  $s_t = \gamma \frac{Y_t}{\ell_{t-1} b_{t-1}} + (1 - \gamma)s_{t-m}$
- Forecast:  $\hat{Z}_{t+h|t} = (\ell_t + b_t h)s_{t-m+h}$

#### 2) Additive Seasonality

- Level:  $\ell_t = \alpha(Z_t - s_{t-m}) + (1 - \alpha)(\ell_{t-1} + b_{t-1})$
- Trend:  $b_t = \beta^*(\ell_t - \ell_{t-1}) + (1 - \beta^*)b_{t-1}$
- Seasonal:  $s_t = \gamma(Z_t - \ell_{t-1} - b_{t-1}) + (1 - \gamma)s_{t-m}$
- Forecast:  $Z_{t+h|t} = (\ell_t + b_t h)s_{t-m+h}$

where  $m$  is the seasonal length.

### C. Combination Method

Forecast combination method is a combination of forecasting of two or more results from individual forecast method. The essence of this method is to give a weight of forecast result as follow

$$f_{ct} = \sum_{j=1}^n w_{jt} f_{jt} = w_{1t} f_{1t} + w_{2t} f_{2t} + \dots + w_{gt} f_{gt}$$

where  $f_{ct}$  is the result of forecast combinations. In this research, there are two combination methods being used that is Simple Average Method combine to individual forecast result with combination forecast equation can be written as follow

$$f_{ct} = \sum_{j=1}^n \frac{f_{jt}}{g}$$

The second method is Variance-Covariance method which one of the best forecast combination method to combine forecast result in optimal manner with the weight system that can be written as follow

$$f_{ct} = w_{1t} f_{1t} + (1 - w_{1t}) f_{2t}$$

In this study, the weight calculation will be modified become Variance-covariance method with shifting window (SWVC). This is done through  $e_{jT}$ , error of forecast for  $j^{th}$  individual models at  $t$ , with weight can be written as follow

$$w_{1t} = \frac{\sum_{T=t-12}^{t-1} e_{2T}^2}{\sum_{T=t-12}^{t-1} e_{1T}^2 + \sum_{T=t-12}^{t-1} e_{2T}^2}$$

with weight change through time. In this study  $m = 12$  is the length of seasonal. The second modification is Seasonal Variance-Covariance Method (SVC),

$$w_{1t} = \frac{e_{2,t-12}^2}{e_{1,t-12}^2 + e_{2,t-12}^2}$$

and for the third is Variance-Covariance Method with fixed Windows. With the weight can be written as follow

$$w_{1t} = \frac{\sum_{T=\xi-12}^{\xi-1} e_{2T}^2}{\sum_{T=\xi-12}^{\xi-1} e_{1T}^2 + \sum_{T=\xi-12}^{\xi-1} e_{2T}^2}$$

### D. Model Selection Criteria

Calculation for the best model is used minimum error of forecast by computing RMSE (Root Mean Square Error) value [3]. The lowest value obtained from this formula is the best model.

$$RMSE = \sqrt{\frac{1}{M} \sum_{t=1}^M e_t^2}$$

## III. METHODOLOGY

The data used in this research is secondary from PT Angkasa Pura I database consisting of the number of passengers not only arrival but also departure at Adisutjipto International Airport, Yogyakarta. Variable used for modelling is monthly number of passengers at arrival and departure terminal for domestic and international flights. The data collected from January 2010 to December 2015 except Bandung route which starts from February 2010 to December 2015. The flight route at Adisutjipto Airport are Denpasar, Pontianak, Balikpapan, Banjarmasin, Bandung, Batam, Jakarta, Ujung Pandang, Surabaya, Kuala Lumpur and Singapore. The step of analysis in this research is as follow.

- 1) Divide the data into two parts, namely in-sample and out-sample
- 2) Forecast using ARIMA method for in-sample data
- 3) Forecast using Exponential Smoothing method for in-sample data
- 4) Forecast using Combination method for in-sample data
- 5) Choose the best forecast method
- 6) Perform forecast with all data (in-sample plus out-sample data)
- 7) Conclusions

## IV. RESULTS AND DISCUSSIONS

Exploration Data in this research is being used to check the growth of number of passengers at the Adisutjipto International Airport Yogyakarta. In-sample data is from 2010 to 2014 period, and remaining data is for out-sample.

Passenger amount for departure and arrival has same pattern where Jakarta is the biggest number and Bandung is the smallest traffics from and to Jogjakarta. It is reasonable because Jakarta is number one in capacity and capability around Indonesia. Bandung has the smallest traffics in case 2010 is year of operational, the most early opened than another airport.

At Adisutjipto International Airport Yogyakarta there are 11 flight routes, consist of 9 domestic flight routes and 2 international flight routes. This paper will only discuss Denpasar flight route which than with the same procedure will be done for the other routes. In Fig. 1, it is shown Denpasar route not only arrival and departure there is growth of the number of passenger every year except at 2015 which shown that the number of passengers who takes flight at Adisutjipto Internal route is going down.

### A. ARIMA Model

The first step in ARIMA method is dividing the data into two parts, namely in-sample and out-sample data using Time

TABLE I: Descriptive Data

City	In-sample		Out-sample	
	Mean	Std. Dev	Mean	Std. Dev
<b>Departure</b>				
Denpasar	21259.00	6097.61	26134.50	3579.31
Bandung	2124.37	756.32	2760.50	546.60
Jakarta	114422.30	22538.00	125327.75	15244.56
Surabaya	10399.33	2397.21	11817.08	2294.29
Balikpapan	19259.42	5131.31	24480.58	3204.22
Banjarmasin	8492.82	3249.59	13513.25	1574.59
Pontianak	4536.53	2244.97	7614.17	1390.46
Ujung Pandang	5744.97	2255.10	10087.17	1654.08
Batam	5099.23	1997.17	8114.00	1391.17
Kuala Lumpur	5361.27	1703.10	7225.92	1183.62
Singapura	4702.47	1279.31	5154.50	713.99
<b>Arrival</b>				
Denpasar	20634.27	5750.82	24755.33	2946.24
Bandung	2128.46	784.24	2985.17	565.52
Jakarta	116695.73	22282.87	127600.08	17202.48
Surabaya	10234.48	2424.53	11102.00	1913.48
Balikpapan	19512.42	5238.60	25400.08	2634.57
Banjarmasin	8637.93	3236.87	13889.33	1381.57
Pontianak	4568.67	2326.14	7633.17	1413.04
Ujung Pandang	6006.90	2447.51	10800.00	1173.77
Batam	4956.98	1778.86	8406.50	1629.82
Kuala Lumpur	6354.07	1878.96	8366.08	1099.53
Singapura	4843.02	1414.80	5444.33	830.035

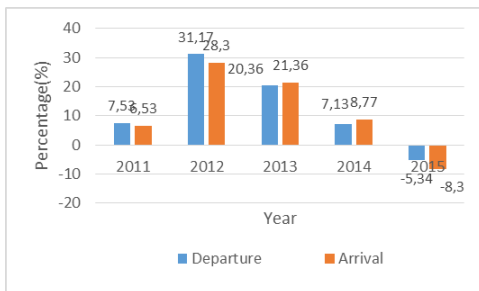


Fig. 1: Percentage of Passengers Growth at Denpasar Route.

Series Plot. From time series plot shows growth from early data to 2015. It can be concluded that in sample data will start from 2010 until 2014 and out sample data will be data for the entire 2015. Second step is checking stationarity of the data. The result shown that both of the data already stationer for variance but not stationer for means. Because of that there is a need for the data to be differenced.

After differencing process is done for one time the data already stationer so the data is ready for the next step of research. Next is identification process using ACF and PACF plot. The ACF plot forming Cut Off pattern but PACF plot forming Tail Off pattern. From the result it is also shown that only lag 1 significant not only ACF but also PACF plot, furthermore the biggest probability of model formed is ARIMA (0,1,1). According to the plot lag 12 is also have influence to the model although not significance. It is shown that there is seasonal factor in this data.

TABLE II: Estimation and Diagnostic Checking

Model	Parameter	Estimation	White noise	Normal distribution
<b>Departure</b>				
ARIMA (0,1,1)	$\theta_t$	0.48299	White noise	No
ARIMA (0,1,1)(1,0,0) <sup>12</sup>	$\Phi_t$	0.65323 0.64554	White Noise	No
<b>Arrival</b>				
ARIMA (0,1,1)	$\theta_t$	0.57757	White noise	No
ARIMA (0,1,1)(1,0,0) <sup>12</sup>	$\Phi_t$	0.65828 -0.48197	White Noise	No

TABLE III: Level, Trend and Seasonal Smoothing Constant

Smoothing	$\alpha$	$\beta$	$\gamma$
<b>Departure</b>			
Holt's Linear Method	0.6481	0.0646	
Holt Winter's Method	0.3544	0.0124	0.0001
<b>Arrival</b>			
Holt's Linear Method	0.5721	0.0975	
Holt Winter's Method	0.3446	0.0151	0.0001

Based on ARIMA models that have been formed, all of parameter are significant in case p-value less than  $\alpha$  (0.05). As for the diagnostic checking for the residuals the data used has met the assumption of White Noise but not for the assumption of Normal distribution. After that if there is more than one of model formed, it must be chosen the best model using RMSE.

### B. Exponential Model

The next method is using Exponential Smoothing. The first step is to identify the most suitable Exponential Smoothing Method to forecast. Number of passenger was identified having trend but it still confusing if the data having a seasonal factor or not. So the method used is Holts Linier Method and Holt Winters Method. The next step is to find the smoothing constant of level, trend and seasonal from the data. The smoothing constant can be seen in the Table III.

Table III shows smoothing constant for the exponential models. It can be seen that for smoothing constant level ( $\alpha$ ) and trend ( $\beta$ ) at Holts Linear Method and Holt Winters Method is different. It happens from the influence of smoothing constant seasonal ( $\gamma$ ). This smoothing constant than can be used to make exponential smoothing model.

### C. Combination Model

The next method is giving a weight to the result of two different individual forecast. The weight used is Simple Average and Variance-Covariance. The weight calculation of Variance-Covariance method uses three different formulas that are Variance-Covariance with Shifting Window (SWVC), Seasonal Varians-Covariance (SVC) dan Variance-Covariance with Fixed Window (FWVC). Three weights obtained from Variance-Covariance and one from Simple Average (SA) then

TABLE IV: RMSE of Combination Model

Route	SA	SWVC	SVC	FWVC
<b>Without seasonal</b>				
Departure	6730.90	<b>6325.13</b>	6522.92	6637.16
Arrival	6323.75	<b>5900.78</b>	6213.63	6200.39
<b>With seasonal</b>				
Departure	4229.07	<b>4110.77</b>	4370.45	4328.65
Arrival	4799.95	<b>4675.73</b>	5120.41	5112.16

TABLE V: RMSE among three models.

Route	ARIMA	Exponential Smoothing	Combination
<b>Without seasonal</b>			
Departure	<b>5275.382</b>	8511.140	6325.133
Arrival	<b>4886.182</b>	7968.861	5900.777
<b>With seasonal</b>			
Departure	<b>3625.786</b>	5025.544	4110.77
Arrival	<b>4036.014</b>	5691.798	4675.73

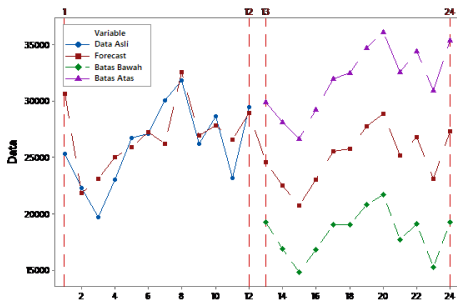


Fig. 2: Plot of Forecasting Departure Passengers.

choose the best combination method. The result can be seen at Table IV.

The best model of combination method is Variance-Covariance with Shifting Windows (SWVC) that has the smallest RMSE.

*D. Choosing The Best Method*

The best method to forecast the number is using RMSE and the results can be seen at Table V.

Seasonal ARIMA is the best model to forecast the number of passengers because the value of RMSE is the lowest than the other method. In this research, modeling with more complexity do not guarantee to give a better result.

*E. Forecast Results*

The next step is to forecast all of the data at that route. The visualization of forecast at can be seen at Fig. 2.

Prediction of passenger has seasonal term with same pattern like this : the first three month (Jan-March) decrease, afterward increase until August. September to November decrease again and rise up on end year December. This period people has Christmas holiday and new year.

With the same method, the best model for every route can be seen at Table VI.

TABLE VI: The Best Model of Route.

Route	Departure	Arrival
Denpasar	ARIMA(0,1,1)(1,0,0) <sup>12</sup>	ARIMA(0,1,1)(0,0,1) <sup>12</sup>
Bandung	*ARIMA([1,5],1,0)	*ARIMA(0,1,1)
Jakarta	ARIMA(0,1,1)(0,0,1) <sup>12</sup>	ARIMA(0,1,1)(1,0,1) <sup>12</sup>
Surabaya	HW	SA
Balikpapan	ARIMA(0,1,1)(0,0,1) <sup>12</sup>	ARIMA(0,1,1)(1,0,0) <sup>12</sup>
Banjarmasin	ARIMA(0,1,1)(1,0,1) <sup>12</sup>	SWVC
Pontianak	FWVC	SVC
Ujung Pandang	HW	SVC
Batam	ARIMA(0,1,1)(1,0,0) <sup>12</sup>	*HLM
Kuala Lumpur	ARIMA(0,1,1)(1,0,0) <sup>12</sup>	SA
Singapura	SWVC	HW

The result of forecast for every departure and arrival route at Adisutjipto International Airport, to get forecast for domestic and international flight at 2016 it is only need to sum all of the forecast result for every route.

V. CONCLUSIONS

The characteristic of flight passengers shows that every route of flight have growth notably for outside Java flight with route that has the most number of passengers is Jakarta route while the least is Bandung route. For all of the analyzed data, it turns out that the best model is the model that incorporate seasonal factor in its analysis except arrival and departure route for Bandung and arrival route for Batam. While the best method to forecast the data of number of passengers is ARIMA or Exponential Smoothing because the result showed that the combination models did not always provide better forecasting result compared to the result of forecasting using individual method.

REFERENCES

- [1] C. Bougas, "Forecasting air passenger traffic flows in canada: an evaluation of time series models and combination methods," 2013.
- [2] S. Makridakis and M. Hibon, "The m3-competition: results, conclusions and implications," *International journal of forecasting*, vol. 16, no. 4, pp. 451–476, 2000.
- [3] A. Chuang, "Time series analysis: univariate and multivariate methods," 1991.
- [4] H. Ravinder, "Forecasting with exponential smoothing whats the right smoothing constant?" *Review of Business Information Systems (RBIS)*, vol. 17, p. 117, 08 2013.
- [5] R. Hyndman, A. Koehler, J. Ord, and R. Snyder, *Forecasting with exponential smoothing: the state space approach*. Springer Science & Business Media, 2008.