Indonesian Journal of Pure and Applied Mathematics
http://journal.uinjkt.ac.id/index.php/inprime
Vol. 1, No. 1 (2019), pp. 57-67
ISSN 2686-5335
doi: 10.15408/inprime.v1i1.12786

Prediction of the Number of Ship Passengers In The Port of Makassar using ARIMAX Method In The Presence of Calendar Variation

Laili Nahlul Farih, Irma Fuziah, and Madona Wijaya
Department of Mathematics, Faculty of Sciences and Technology, UIN Syarif Hidayatullah Jakarta
Email: lailinahlul@gmail.com

Abstract

Indonesia is an archipelago with the largest Muslim population in the world. Every year, Indonesian people have a tradition of meeting relatives in other areas or take a vacation on Eid al-Fitr. People use different modes of transport to travel such as air, water, and land transport. Port plays a role in supporting water transportation because it is a knot of inter-regional relations. The celebration of Eid al-Fitr moves forward by about 11 days every year. The purpose of this thesis is to make an estimate of the total departure of ship passengers in the main port of Makassar using the ARIMAX method with the effects of calendar variations. The ARIMAX method is a method that can be used when there are exogenous variables, where in this case the exogenous variable is in the form of variable dummy wich is Eid holidays. These forecasting results show that the ARIMAX $(2,1,0)(0,0,1)^{12}$ method has a relatively small accuracy with the MAPE value of 14.08%.
Keywords: water transportation; calendar variations effects; Eid Al-Fitr.

Abstrak

Indonesia merupakan negara kepulauan dengan mayoritas muslim terbesar didunia. Setiap tahun masyarakat Indonesia memiliki tradisi bertemu sanak saudara di daerah lain ataupun berlibur pada hari raya Idul Fitri. Jalur transportasi yang digunakan yaitu melalui darat, udara dan laut. Pelabuhan memiliki peran yang sangat penting dalam mendukung transportasi laut karena menjadi titik simpul hubungan antar daerah. Perayaan hari raya Idul Fitri dalam setiap tahun mengalami pergeseran 11 hari. Tujuan penulisan skripsi ini adalah untuk membuat prakiraan total keberangkatan penumpang kapal di Pelabuhan Utama Makassar menggunakan metode ARIMAX dengan efek variasi kalender. Metode ARIMAX merupakan metode yang dapat digunakan ketika data tersebut menggunakan variable eksogen, dimana dalam kasus ini variable eksogennya berupa variable *dummy* libur hari raya idul fitri. Hasil peramalan ini menunjukan bahwa metode ARIMAX $(2,1,0)(0,0,1)^{12}$ memiliki tingkat akurasi yang lebih baik dibandingkan ARIMA musiman $(2,1,0)(0,0,1)^{12}$ dengan nilai MAPE sebesar 14,08%.
Kata Kunci: transportasi air; efek variasi kalender, Hari Raya Idul Fitri.

1. INTRODUCTION

Indonesia is the largest archipelago country consisting of 17,504 islands with a population of over 270 million people and is one of the largest Muslim population in the world with more than 230 million people [1]. On Eid al-Fitr holidays, Indonesian people make this important moment for homecoming trip, going on vacation, or meeting relatives who live in other areas. Besides using air and land transportation, water transportation is also one of many other choices for people to travel.

Every year, Eid al-Fitr celebration is determined based on the Islamic calendar (Hijriyah) which is 11 days shorter than the Gregorian year. This implies that the days of Eid al-Fitr always change each year. This changes in Georgian calendar causes the number of transportations that is used for homecoming trips increases in different months each year. Different patterns for time series data for total departing passengers in main port of Makassar affect the prediction number of departing passengers in main port of Makassar.

Many research studies in the effect of calendar variations have been conducted, it was initiated by Liu [2]. Liu [2] studied the effect of holidays using identification and estimation of ARIMA model on monthly traffic data in Taiwan. Cleveland and Devlin [3] examined methods to overcome the effect of calendar variation on economic data in the United States. Hillmer [4] described about time series model on telephone data by adding variation of daily trading. Nasiru et.al [5] discussed about currency circulation in Ghana. Nasiru et al. [5] explained about the effect of calendar variation on Eid al-Fitr days on children's clothing sales using ARIMAX method. In this research, we will model and predict the number of departing ship's passengers over month by incorporating the effect of calendar variation in the model.

2. METHOD

2.1. Regression Model with Dummy Variables

Regression model with dummy variables is a method that quantifies a categorical predictor variable by creating artificial or dummy variables to represent an attribute with two or more categories and include them in the regression coefficient by assigning a value of 1 or 0. A value of 1 indicates the presence of some categories and 0 indicates the absence of some categories. A regression model with dummy variables can be expressed as follows [6]:

$$\boldsymbol{Y_t = \beta_0 + \beta_1 V_{1,t} + \beta_2 V_{2,t} + \cdots + \beta_m V_{i,t} + w_t}, \tag{1}$$

where β_0 is the intercept, $\beta_1, \beta_2, \ldots, \beta_m$ are the parameter coefficients (slopes) and w_t is the residual

2.2. Stationarity

The important assumption that should be satisfied in time series model is stationarity. The basic idea of stationarity is that the probability law that governs the behavior of the process does not change with time. Assumption that need to be fulfilled in ARIMA model is (weak) stationarity [7]. Time series model is said a stationary series if:

a. $\mu_t = E[y_t] = \mu$, meaning the mean function is constant over time, and
b. $\gamma_{t,t-k} = \gamma_{0,k}$, meaning the covariance function depends only the time difference (lag) and not an the actual time.

2.3. Augmented Dickey-Fuller Test (ADF)

Unit root test is used to test if a series is stationary. It is performed by evaluating if a unit root is present in a time series sample. An ADF tests the null hypothesis that a unit root is present in sample (H_0: $\hat{\delta} = 0$) versus H_1: $\hat{\delta} < 0$ with the test statistic:

$$\textbf{ADF } \boldsymbol{test} = \frac{\hat{\boldsymbol{\delta}}}{\boldsymbol{\sigma}(\hat{\boldsymbol{\delta}})}, \tag{2}$$

where $\hat{\delta}$ is the estimated least square of δ and $\sigma(\hat{\delta})$ is the standard error from the estimated least square of $\hat{\delta}$. We reject the null hypothesis if the p-value is less than the significance level (α =5%).

2.4. Box-Cox Transformation

A stationary series means that it has a constant mean and variance. If the data is not yet stationary in variance then a transformation to the sample data needs to be conducted. One of the most famous methods for data transformation is called Box-Cox transformation. In general, Box-Cox transformation can be expressed as in the followings [10]:

$$\boldsymbol{T}(\boldsymbol{Y_t}) = \frac{\boldsymbol{Y_t}^{\boldsymbol{\lambda}} - \mathbf{1}}{\boldsymbol{\lambda}}. \tag{3}$$

2.5. SARIMA Model

SARIMA model is an ARIMA model with seasonality effects. The seasonal part of the model consists of terms that are similar to the non-seasonal components of the model, but involve backshifts of the seasonal period and can be expressed as follows [10]:

$$\boldsymbol{\Phi_p}(\boldsymbol{B^s})\boldsymbol{\varphi_p}(\boldsymbol{B})(\mathbf{1}-\boldsymbol{B})^{\boldsymbol{d}}(\mathbf{1}-\boldsymbol{B^s})^{\boldsymbol{D}}\boldsymbol{Y_t} = \boldsymbol{\theta_q}(\boldsymbol{B})\boldsymbol{\Theta_Q}(\boldsymbol{B^s})\boldsymbol{e_t}. \tag{4}$$

Such models are expressed as *(p, d, q)* × *(P, D, Q)s.* where *(p, d, q)* are as for an ARIMA model, while *(P, D, Q)m* express the seasonal autoregressive, integration and moving average components where the seasonality period is *s. H*ere, we have *P* seasonal autoregressive terms (with coefficients Φ_1, …, Φ_P), *Q* seasonal moving average terms (with coefficients Θ_1, …, Θ_Q) and *D* seasonal differencing based on *m* seasonal periods.

2.6. The ARIMA Procedure

Box-Jenkins method in fitting ARIMA model generally consists of 4 steps [9]:
The first step in developing a Box-Jenkins model is model identification. The choice of the model can be seen from autocorrelation function (ACF) and partial autocorrelation function (PACF) [9]. ACF is a function that quantifies the correlation between observation at time t and the observations at previous time. Meanwhile, PACF is a function that quantifies partial correlation between the series and lags of itself, i.e. correlation between them which is not explained by their mutual correlations with a specified set of other variables [10]. To determine the order of ARMA model, we can observe from the pattern of ACF and PACF as summarized in Table 1 [9].

Table 1. Behavior of the ACF and PACF for ARMA model

Proses	ACF	PACF
AR (p)	Tails off	Cuts off after lag p
MA (q)	Cuts off after lag q	Tails off
ARMA (p, q)	Tails off	Tails off

The second step is parameter estimation of ARIMA model using Least Square method. This method has several assumptions that need to be fulfilled, where residuals e_t have to have 0 mean ($E(e_t) = 0$), variance of square of residuals is constant $E({e_t}^2) = {\sigma_e}^2$, and no correlation between residuals ($E(e_t e_k) = 0$) for $t \neq k$ and does not correlate with explanatory variable V_t i.e. $E(V_t e_t) = 0$ [10]. The coefficient of parameters from the estimated model are then tested. The null hypothesis is the coefficients are not significant [11] with test statistics is:

$$t = \frac{\hat{\theta}}{se(\hat{\theta})}$$

The null hypothesis is rejected if the p-values is less than $\alpha = 5\%$.

The next step is model diagnostic and model selection. In diagnostic test, the best fitted model is selected by looking if the model residuals are white noise, i.e. have constant mean and variance, and no correlation between residuals [9]. The residual analyses are performed based on the following [7]:

1. Plot of Residual

The first step in model diagnostic is by plotting the residual over time t. If the model fits the data well, then the residuals will show no pattern will and the points will scatter within the horizontal lines [7]. This means, the residual plot from a model is used to see if the mean and variance of the model residuals are constant or not.

2. Test the autocorrelation of the residuals

The second step to check if the residuals are white noise is by evaluating the ACF of the residuals. If the model fits the data well, the sample residuals should not correlated to each other. The hypothesis for Ljung-Box test is as follows [9]:

$H_0 : \rho_1 = \rho_2 = \cdots = \rho_k = 0$ (no correlation between the series of residual up to lag k, $k < n$),
H_1: at least $\rho_k \neq 0$, for $k = 1,2,\dots,n$.

The test statistic is $Q = n(n+2)\frac{\sum_{j=1}^{k}\hat{\rho}(j)^2}{(n-j)}$ follows $\chi^2\big(k-(p+q)\big), k > (p+q)$ where k is the maximum of lag length, n is the total data sample, $\hat{\rho}(j)$ is the ACF from residual at lag j. The criteria to reject H_0 if $Q > \chi^2_{tabel}$ or if $p-value < \alpha = 0.05$ with significance level of 5%.

After the best fitted model is obtained from the modelling steps above, the model is then used to forecast the data in the future. In time series analysis, dataset can be divided into two parts, i.e. in-sample and out-sample data. In-sample data is used to select the best model meanwhile out-sample data is used to validate the accuracy of forecasting from the selected model obtained from fitting to in-sample data [9].

Two criteria that can be used to test the accuracy of the model such as [9]:

1. Mean Squared Error (MSE)

$$\mathbf{MSE} = \frac{\sum_{i=1}^{m}(Y_i-\hat{Y}_i)^2}{n}, \mathbf{m} < \mathbf{n}. \tag{5}$$

where Y_i is actual data, $\widehat{Y_i}$ is the predicted data, and n is the number of data sample.

2. Mean Absolute Percentage Error (MAPE)

$$\mathbf{MAPE} = \frac{100\%}{n}\sum_{t=1}^{n}\left|\frac{Y_i-\hat{Y}_i}{Y_i}\right|. \tag{6}$$

ARIMAX model is an extension of ARIMA model with addition of exogenous variables. This model is a combination of a regression model and ARIMA model [5]. The development of ARIMA model is started from equation (1) with w_t is the regression residuals. If w_t is white noise then (1) is called a regression model, but if w_t is not white noise then it is seasonal ARIMA model $(p,d,q)(P,D,Q)^S$ as denoted in equation (4).

Exogenous variable used in this study is a dummy variable of calendar variation effect. Calendar variation is a seasonal trend with variation of length of period. Calendar variation can be caused by variation in school holidays or other major religious holidays from month to month or from year to year [6]. The equation of ARIMA model with calendar variation effect is a combination equation (1) and (4) and this can be expressed as follows [11]:

$$Y_t = \beta_0 + \beta_1, V_{1,t} + \cdots + \beta_m, V_{i,t} + \frac{\theta_q(B)\Theta_Q(B^S)}{\phi_p(B)\Phi_P(B^S)(1-B)^d(1-B^S)^D} e_t \quad (7)$$

where $V_{1,t}, V_{2,t}, \ldots, V_{i,t}$ are the dummy variables; $\beta_{s1}, \beta_{s2}, \beta_{s(L-1)}$ are the parameter of dummy variables, $(1-B)^d$ is the differencing operator of order d, and $(1-B^S)^D$ is the seasonal differencing operator of order D.

2.7. The Data

The data used in this study is a secondary dataset of total departing ship's passengers in main port of Makassar in period of January 2006 to December 2017. The data is collected on a monthly basis resulting with 144 observations. The data is obtained from Central Bureau of Statistics (BPS) [12] and is then divided into two parts with percentage of 80% and 20% for model building and model validation, respectively. The data obtained in period January 2006 to December 2015 is used for model building (observation data) and the remaining from January 2016 to December 2017 is used for model validation.

There are two variables used in this study i.e. a response variable (the number of departing ship's passengers in main port of Makassar) and predictor variables (the effect of calendar variation considered as dummy variables). The dummy variable, V_t, with 1 represent a month one week before and after Eid al-Fitr holiday and 0 represents other months.

This study utilizes RStudio 3.5.1 software to analyze the number of departing ship's passengers in main port of Makassar. The steps to analyze the data are shown in the following:

1. Summarizing the descriptive statistics of the data and determining the dummy variables to represent the effect of calendar variation.
2. Fitting regression model with the effect of calendar variation as the predictor (equation (1)).
3. Testing the stationarity of regression residuals obtained in point 2 with Box-Cox test to evaluate the stationarity in variance and ADF test to evaluate the stationarity in means.
4. Performing diagnostic test to examine if the regression residuals are white noise with Ljung-Box test.
5. If the residuals do not follow white nose process then an ARIMA model is identified from the residual series by plotting the ACF and PACF.
6. Fitting ARIMAX model by estimating the model in point 2 and 5 simultaneously.
   a. Performing model diagnosis to the fitted ARIMAX model: create plot of residuals obtained from the selected ARIMAX model to see if there is a trend or not, and then test with Ljung-Box test on model residuals to investigate if the assumption of white noise is met or not.
7. Forecasting.

a. Predicting the number of departing ship's passengers in main port of Makassar by including the effect of calendar variation.
b. Computing the accuracy of the forecasted data to show how close the predicted data to the actual data using MSE and MAPE.
c. Forecasting the data for period of January 2017 to December 2017 by using ARIMAX model with the smallest MSE and MAPE values.

3. RESULTS AND DISCUSSIONS

The number of departing ship's passengers in main port of Makassar on average are 40,393 passengers every month with standard deviation 14,509. The lowest number of passengers was 18,191 and the highest reach 117,195 passengers.

The effect of calendar variation used in this study is one week before and after the Eid al-Fitr holidays. The first step carried out is by fitting regression model to the data with the effect of calendar variation as dummy variables. The fitted regression equation with dummy variable of the calendar variation effect can be written as follows:

$$Y_t = 37227 + 27138v_t + w_t,$$

where v_t represents the dummy variable of one week before Eid al-Fitr holiday and w_t is the regression residuals. The effect of calendar variation was found significant since the p-value is less than 0.05 (p=0.0005). The regression residuals are then tested with Ljung Box test. If the residuals is found insignificant then there is no need to further conduct the analysis with SARIMA since regression model alone is enough to model the result suggests that null hypothesis of no correlation was rejected at 5% level since p-value = 0.0003 and is less than 0.05. Thus we conclude that the residuals w_t are correlated. Since the residuals are correlated then we fit an ARIMA model and identify the orders of model. But before that, we transform the series using Box-Cox transformation to ensure that the residuals are stationary in variance and also perform ADF test to check if the residuals are stationary in mean. Figure 1 (left) suggests that the data are not yet stationary in variance because the rounded value does not approach 1, and also the lower and upper interval do not equal to 1. According to the estimated lambda, the series need to be transformed using a square root, i.e. $\sqrt{w_t}$. The ADF test was carried out the transformed series, the results indicates the data is not stationary in means because we fail to reject the null hypothesis, the p-value is 0.6753 and is greater than 0.05. Since the data is not stationary in means, then we perform first difference to the data to help stabilize the mean. ADF test is again carried out to the first differencing data and the result shows that the series exhibit stationary in the first difference since the p-value is less than 0.05 (*p value* = 0.01). Figure 1 (right) displays the transformed series after applying square root to the data as well as first differencing.

Figure 2 displays the ACF and PACF plots. They show that the ACF plot cuts off at lag 1, 11, and 12. The "cut off" at lag 11 and 12 indicates that there might be a seasonal effect. The PACF plot shows cut-off at lag 1, 2, 4, 5, 10, 11, and 12. The "cut off" at lag 4, 5, 10, 11, and 12 might be an indication of seasonal effect. According to these plots, the candidates for ARIMA models are ARIMA $(1,1,1)(1,0,1)^{12}$, ARIMA $(1,1,1)(1,0,0)^{12}$, ARIMA $(1,1,1)(0,0,1)^{12}$, ARIMA $(1,1,0)(1,0,1)^{12}$, ARIMA $(1,1,0)(1,0,0)^{12}$, ARIMA $(1,1,0)(0,0,1)^{12}$, ARIMA $(0,1,1)(1,0,1)^{12}$, ARIMA $(0,1,1)(1,0,0)^{12}$, ARIMA $(0,1,1)(0,0,1)^{12}$, ARIMA $(2,1,1)(1,0,1)^{12}$, ARIMA $(2,1,1)(1,0,0)^{12}$, ARIMA $(2,1,1)(0,0,1)^{12}$, RIMA $(2,1,0)(1,0,1)^{12}$, ARIMA $(2,1,0)(1,0,0)^{12}$ and ARIMA $(2,1,0)(0,0,1)^{12}$.

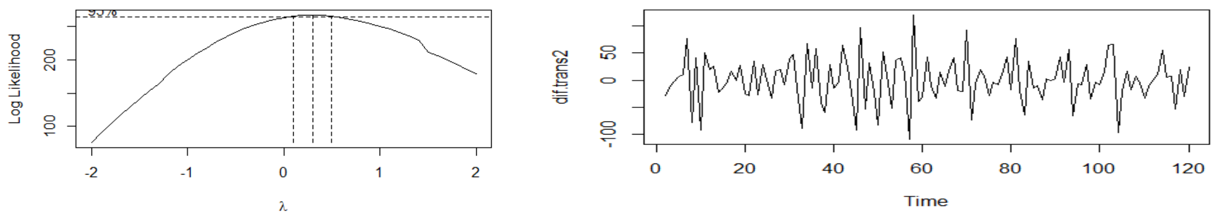

Figure 1. Box-Cox transformation for the regression residuals (left), and plot of first difference of the square root of regression (right)

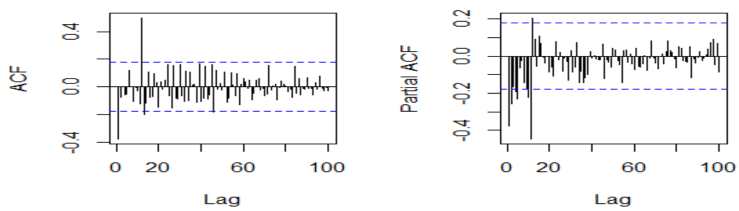

Figure 2. Plot of ACF and PACF

After ARIMA model is fitted to the w_t series, we fit ARIMAX model by estimating the regression model with calendar variation effect and w_t series that was fitted with ARIMA model simultaneously. The estimated parameters from the 15 models show that ARIMAX $(1,1,1)(1,0,1)^{12}$, ARIMAX $(1,1,0)(1,0,1)^{12}$, ARIMAX $(0,1,1)(1,0,1)^{12}$, ARIMAX $(2,1,1)(1,0,1)^{12}$, ARIMAX $(2,1,1)(1,0,0)^{12}$, ARIMAX $(2,1,1)(0,0,1)^{12}$, and ARIMAX $(2,1,0)(1,0,1)^{12}$ have insignificant parameters, thus the models that are tested in the next step in the diagnostic models are ARIMAX $(1,1,1)(1,0,0)^{12}$, ARIMAX $(1,1,1)(0,0,1)^{12}$, ARIMAX $(1,1,0)(1,0,0)^{12}$, ARIMAX $(0,1,1)(0,0,1)^{12}$, ARIMAX $(0,1,1)(1,0,0)^{12}$, ARIMAX $(1,1,0)(0,0,1)^{12}$, ARIMAX $(2,1,0)(1,0,0)^{12}$ and ARIMAX $(2,1,0)(0,0,1)^{12}$. ARIMAX model with significant parameters can be seen in Table 2.

Before computing the forecasting values in the future, it is important to check the assumption of residuals from fitting ARIMA model. The assumption for the model residuals are stationarity and white noise. The residual plot is firstly examined as shown in Figure 3. The plot indicates that there is no obvious trend. The second step is by testing the correlation of the residuals by using Ljung-Box test. Table 3 suggests that the results of autocorrelation test for ARIMAX$(1,1,1)(1,0,0)^{12}$, ARIMAX$(1,1,1)(0,0,1)^{12}$, ARIMAX$(1,1,0)(1,0,0)^{12}$, ARIMAX$(1,1,0)(0,0,1)^{12}$, ARIMAX $(0,1,1)(1,0,0)^{12}$, ARIMAX$(0,1,1)(0,0,1)^{12}$, ARIMAX$(2,1,0)(1,0,0)^{12}$ and ARIMAX $(2,1,0)(0,0,1)^{12}$ have p-value greater than 0.05, and thus the null hypothesis is rejecting. This means that the residuals are not correlated.

After model diagnostics indicate that the model assumptions are met, then validation of forecasting the number of departing passengers in main port of Makassar is carrier out. In this study, MSE and MAPE are used to measure the forecasting accuracy. Table 4 shows that the smallest values of MSE and MAPE are obtained from ARIMAX $(2,1,0)(0,0,1)^{12}$ model. The selected model with the smallest MSE and MAPE values is compared with seasonal ARIMA model as shown in Table 5. Figure 4 shows that the predicted number of departing passengers using ARIMAX $(2,1,0)(0,0,1)^{12}$ is relatively close to the actual data, meanwhile the seasonal ARIMA $(2,1,0)(0,0,1)^{12}$ model as depicted in Figure 5 shows that the predicted number of ship's passengers are still far away. Table 4 also indicates that ARIMAX $(2,1,0)(0,0,1)^{12}$ model is superior to seasonal ARIMA $(2,1,0)(0,0,1)^{12}$ model.

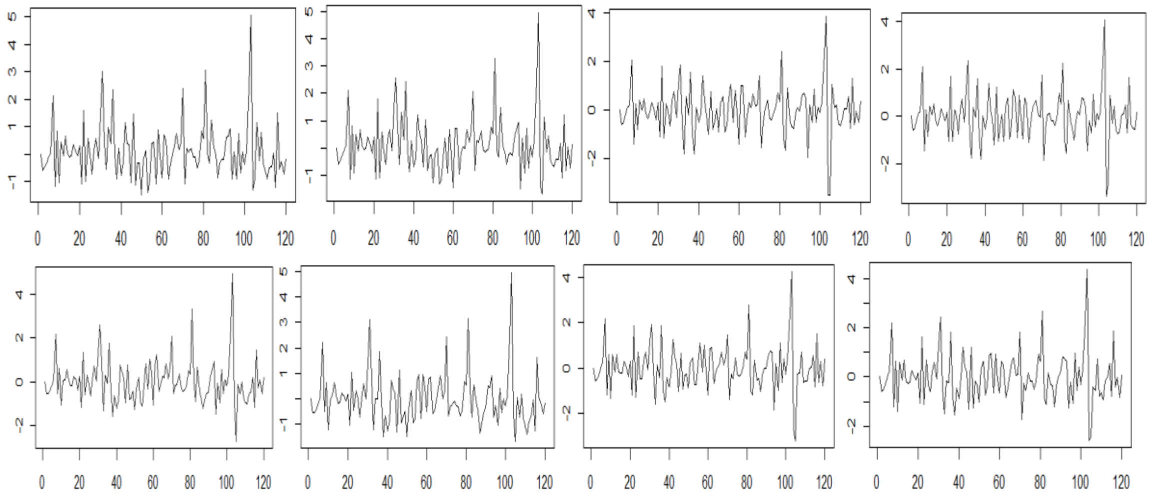

Figure 3. Plot of model residuals of ARIMAX $(\mathbf{1,1,1})(\mathbf{1,0,0})^{\mathbf{12}}$, ARIMAX $(\mathbf{1,1,1})(\mathbf{0,0,1})^{\mathbf{12}}$, ARIMAX $(\mathbf{1,1,0})(\mathbf{1,0,0})^{\mathbf{12}}$, ARIMAX $(\mathbf{1,1,0})(\mathbf{0,0,1})^{\mathbf{12}}$, ARIMAX $(\mathbf{0,1,1})(\mathbf{1,0,0})^{\mathbf{12}}$, ARIMAX $(\mathbf{0,1,1})(\mathbf{0,0,1})^{\mathbf{12}}$, ARIMAX $(\mathbf{2,1,0})(\mathbf{1,0,0})^{\mathbf{12}}$, and ARIMAX $(\mathbf{2,1,0})(\mathbf{0,0,1})^{\mathbf{12}}$.

Table 2. The estimated parameters of ARIMAX model

No	Model	Parameter	Estimated Coefficient	Std. Error	Significant
1.	ARIMAX $(1,1,1)(1,0,0)^{12}$	ϕ_1	0.3811	0.0940	Yes
		θ_1	-0.9872	0,0344	Yes
		Φ_1	0.4912	0.0802	Yes
		v_t	20373.91	3514.15	Yes
2.	ARIMAX $(1,1,1)(0,0,1)^{12}$	ϕ_1	0.3760	0.0955	Yes
		θ_1	-0.9881	0.0392	Yes
		Θ_1	0.5320	0.0882	Yes
		v_t	21339.94	3587.89	Yes
3.	ARIMAX $(1,1,0)(1,0,0)^{12}$	ϕ_1	-0.3265	0.0877	Yes
		Φ_1	0.4923	0.0783	Yes
		v_t	19194.08	3239.53	Yes
4.	ARIMAX $(1,1,0)(0,0,1)^{12}$	ϕ_1	-0.3196	0.0878	Yes
		Θ_1	0.5316	0.0874	Yes
		v_t	19536.12	3209.00	Yes
5.	ARIMAX $(0,1,1)(1,0,0)^{12}$	θ_1	-0.7631	0.0952	Yes
		Φ_1	0.5265	0.0776	Yes
		v_t	23292.14	3667.48	Yes
6.	ARIMAX $(0,1,1)(0,0,1)^{12}$	θ_1	-0.7852	0.1152	Yes
		Θ_1	0.5525	0.0886	Yes
		v_t	24983.23	3683.76	Yes
7.	ARIMAX $(2,1,0)(1,0,0)^{12}$	ϕ_1	-0.4195	0.0908	Yes
		ϕ_2	-0.2651	0.0884	Yes
		Φ_1	0.5190	0.0784	Yes
		v_t	19651.70	3332.52	Yes
8.	ARIMAX $(2,1,0)(0,0,1)^{12}$	ϕ_1	-0.4108	0.0912	Yes
		ϕ_2	-0.2581	0.0887	Yes
		Θ_1	0.5547	0.0864	Yes
		v_t	20614.68	3321.37	Yes

Table 3. The results of Ljung-Box test

Model	p-value	Decision
ARIMAX $(1,1,1)(1,0,0)^{12}$	0.7542	Do not reject H_0
ARIMAX $(1,1,1)(0,0,1)^{12}$	0.7817	Do not reject H_0
ARIMAX $(1,1,0)(1,0,0)^{12}$	0.3358	Do not reject H_0
ARIMAX $(1,1,0)(0,0,1)^{12}$	0.3581	Do not reject H_0
ARIMAX $(0,1,1)(1,0,0)^{12}$	0.0963	Do not reject H_0
ARIMAX $(0,1,1)(0,0,1)^{12}$	0.0511	Do not reject H_0
ARIMAX$(2,1,0)(1,0,0)^{12}$	0.4983	Do not reject H_0
ARIMAX $(2,1,0)(0,0,1)^{12}$	0.5824	Do not reject H_0

Table 4. Forecasting accuracy of ARIMAX model with MSE and MAPE

Model	MSE	MAPE
ARIMAX $(1,1,1)(1,0,0)^{12}$	96065572	32.14
ARIMAX $(1,1,1)(0,0,1)^{12}$	92565749	31.83
ARIMAX $(1,1,0)(1,0,0)^{12}$	38353142	19.55
ARIMAX $(1,1,0)(0,0,1)^{12}$	32758315	14.64
ARIMAX $(0,1,1)(1,0,0)^{12}$	43945094	19.76
ARIMAX $(0,1,1)(0,0,1)^{12}$	43358655	19.24
ARIMAX$(2,1,0)(1,0,0)^{12}$	36892444	18.81
ARIMAX $(2,1,0)(0,0,1)^{12}$	31760940	14.08

Table 5. Comparison of forecasting values between ARIMAX $(\mathbf{2},\mathbf{1},\mathbf{0})(\mathbf{0},\mathbf{0},\mathbf{1})^{\mathbf{12}}$ and SARIMA $(\mathbf{2},\mathbf{1},\mathbf{0})(\mathbf{0},\mathbf{0},\mathbf{1})^{\mathbf{12}}$

Period	Actual data	ARIMAX $(2,1,0)(0,0,1)^{12}$	Seasonal ARIMA $(2,1,0)(0,0,1)^{12}$
January 2016	33941	33216.67	33897.90
February 2016	22871	29721.85	30783.60
March 2016	26146	27088.59	28178.30
April 2016	23515	28212.63	29147.50
May 2016	28466	28574.71	29818.48
June 2016	43548	31074.61	32819.71
July 2016	56437	46017.68	36158.90
August 2016	30687	37760.54	35404.79
September 2016	28768	31961.14	32627.51
October 2016	25319	30929.73	32409.30
November 2016	25345	28261.77	29691.94
⋮	⋮	⋮	⋮
November 2017	25345	29672.55	31231.73
December 2017	31146	29672.48	31231.84
MSE		31760940	77803011
MAPE		14,08	19.22

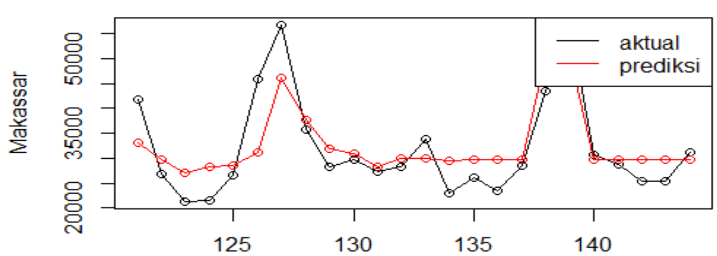

Figure 4. Plot of predicted vs. actual number of departing ship's passengers using ARIMAX$(\mathbf{2},\mathbf{1},\mathbf{0})(\mathbf{0},\mathbf{0},\mathbf{1})^{\mathbf{12}}$

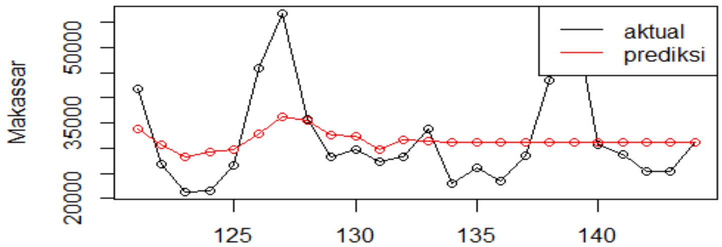

Figure 5. Plot of predicted vs. actual number of departing ship's passengers using seasonal ARIMA$(\mathbf{2},\mathbf{1},\mathbf{0})(\mathbf{0},\mathbf{0},\mathbf{1})^{\mathbf{12}}$

The results of forecasting the number of departing ship's passengers in main port of Makassar in period of January 2018 to December 2018 can be seen in Table 6. The table indicates that the increase in the number of passengers was found in June 2018 because of Eid al-Fit holiday falls in this month.

Table 6. Forecasting results

Period	The forecasting results from ARIMAX $(2,1,0)(0,0,1)^{12}$ model
January 2018	33216.67
February 2018	29721.85
March 2018	27088.59
April 2018	28212.63
May 2018	28574.71
Jun3 2018	51689.29
July 2018	25402.99
August 2018	37760.54
September 2018	31961.14
October 2018	30929.73
November 2018	28261.77
December 2018	29927.81

4. CONCLUSION

According to the result and discussion section, the total of departing ship passengers in Belawan port can be forecasted using the best selected ARIMAX model, i.e. ARIMAX $(2,1,0)(0,0,1)^{12}$ with exogenous variable of dummy variables indicating one week before and after Eid al-Fitr holiday. This model has the smallest MSE value of 31760940 and MAPE of $14,08$. The equation of ARIMAX model can be written as follows:

$$\hat{Y}_t = 17359{,}61\, V_t + \frac{(1-0{,}5547B^{12})}{(1+0{,}4108B+0{,}2581B^2)(1-B)^1} e_t$$

The forecasting results of the number of departing passengers experience a high surge on June 2018 (~ 51689 passengers) because this month is the month of Eid al-Fitr holidays.

We propose for next study to include other exogenous variables that have influence on the number of departing passengers in main port of Makassar and also conduct a comparison study with other methods.

REFERENCES

[1] BPS, "Proyeksi Penduduk Indonesia 2010-2035," The Central Berau of Statistics, Jakarta, 1913.

[2] L.-M. Liu, "Analysis of Time Series with Calendar Effects," *Management Science,* vol. 26, pp. 106-112, 1980.

[3] W. Cleveland and S. Devlin, "Calendar Effects in Monthly Time Series: Modeling and Adjustment," *Journal of The American Statistical Association,* vol. 77, pp. 520-528, 1982.

[4] S. Hillmer, "Forecasting Time Series with Trading Day Variation," *Journal of Forecasting,* vol. 1, pp. 385-395, 1982.

[5] S. Nasiru, A. Luguterah and L. Anzagra, "The Efficacy of ARIMAX and SARIMA Models in Predicting Monthly Currency in Circulation in Ghana," *Mathematical Theory and Modeling,* vol. 3, pp. 73-81, 2013.

[6] A. Widarjono, Ekonometrika Pengantar dan Aplikasinya disertai Panduan EViews, Yogyakarta: UPP STIM YKPN, 2013.

[7] J. D. Cryer and K.-S. Chan, Time Series Analysis with Aplications in R Second Edition, New York: Springer-Verlag, 2008.

[8] W. W. Wei, Time Series Analysis: Univariate and Multivariate Methods. Second Edition, United States of America: Greg Tobin, 2006.

[9] D. Rosadi, Analisis Ekonometrika & Runtun Waktu Terapan dengan R, Yogyakarta: KOMINFO, 2011.

[10] R. S. Tsay, Analysis of Financial Tima Series Second Edition, New York: John Wiley and Sons, Inc, 2005.

[11] M. H. Lee, S. and N. A. Hamzah, "Calendar Variation Model Based on ARIMAX for Forecasting Sales Data with Ramadhan Effect," in *Proceedings on the Regional Conference on Statistical Sciences (RCSS' 10)*, Malaysia, 2010.

[12] B. P. Statistik, "https://www.bps.go.id/," [Online]. Available: https://www.bps.go.id/dynamictable/2015/03/10/818/total-keberangkatan-penumpang-dari-pelayaran-dalam-negeri-di-5-pelabuhan-utama-2006-2018-orang-.html. [Accessed 24 Juli 2018].

[13] L.-M. Liu, "Identification of Time Series Models In The Presence of Calendar Variation," *International Journal of Forecasting,* vol. 2, pp. 357-372, 1986.