Forecasting the Number of Vehicles in Indonesia Using Auto Regressive Integrative Moving Average (ARIMA) Method

Elmir Arif Irhami¹, F. Farizal²

Department of Industrial Engineering, Universitas Indonesia, Depok 16424, Indonesia¹,²

Email: elmir.arif@ui.ac.id¹, farizal@eng.ui.ac.id²

Abstract. The number of vehicles in Indonesia increases significantly from year to year. This has an impact on various aspects such as traffic jams, air pollution, traffic accidents etc. The purpose of this study is to obtain the best model for forecast the number of cars and the number of motorcycles in the next 11 years. For the purpose, ARIMA method was used. Using the historical data of the number of cars and the number of motorcycles from 2001 to 2019, the best model for forecasting the number of cars and the number of motorcycles is ARIMA (1,1,0) and ARIMA (2,1,2), respectively. The models have MAPE of 7.01% and 7.24% for cars and motorcycles, respectively.

Keywords: Number of Vehicle, Forecasting and ARIMA.

1. Introduction

Vehicle is a tool that can facilitate a variety of daily activities, for example transportation to work. Vehicle as a mean of transportation consists of motorized and non-motorized vehicles. Motorized vehicles are vehicles that are driven by a machine and whereas non-motorized vehicles are vehicles that are driven by the power of living things or that are not from a machine. The definition of motorized vehicles is listed in Article 1 paragraph 8 of Law Number 22 Year 2009 concerning Traffic and Transportation (hereinafter referred to as Law Number 22 Year 2009) "Motorized Vehicles are any vehicles that are driven by mechanical equipment in the form of machines other than vehicles that run on rail".

At present there are 3 types of transportation including land transportation, sea transportation and air transportation. Based on data from the Central Statistics Agency (BPS) from January 2019 to June 2019, land transportation is the most commonly used transportation with a number of 209 million users compared to water and air transportation, respectively 11 million people and 36 million people. According to data obtained from the Central Statistics Agency (BPS), nationally the number of domestic land, sea and air transportation passengers from 2016 to 2019 is recorded to have experienced a significant increase.

Many developed countries consider that the development of transportation is integrated with the development of the economy. This is because the development of economic activity is directly proportional to the development of transportation (Salim, 2006). Transportation is commonly called the economic artery because transportation is one of the means that has a role in economic development and regional development. The role of transportation in the economy is as a benchmark in the economic development of an area. Growth in the number of vehicles that is fairly high from year to year has an impact on various aspects such as traffic jams, air pollution and traffic accidents.
Therefore, by forecasting the number of motorized vehicles, it will be able to help the relevant parties to make a policy or program that can reduce the problems that arise due to the high number of vehicles. For example, by forecasting the number of vehicles better, the government can estimate the amount of fuel subsidy more precisely (Farizal et al., 2014). Forecasting is the process of estimating (measuring) the amount or amount of something in the future based on past data that is analyzed scientifically, especially using statistical methods (Sudjana, 1989).

Research by Fajar Pangestu et al. (2018) which predicts motor vehicles in Indonesia using the Average-Based Fuzzy Time Series Model method. In their research, they found that the large amount of data used in the prediction led to an increase in the value of MAPE. This study uses data on cars and motorbikes which are both used as 1 variable (Pangestu, Fajar et al., 2018). Another study by Rizki Febrianto, 2017 using the ARIMA method to predict the number of private cars in Surabaya. The results of forecasting the number of private cars in the city of Surabaya in January 2017 to December 2017 has increased continuously. With an average estimated number of private cars in the city of Surabaya in 2017 amounting to 763343.9 private car users, with an average increase of 2823.36 car vehicles each month in 2017 (Febrianto, 2017). Vaghasia (2018) predicts traffic flow using a hybrid approach that combines ARIMA model with fuzzy wavelet transform. In other paper use ARIMA model top predict automobile sales for the next following 5 years (Shakti, 2017). Other paper use ARIMA model to analyze and predict observations of traffic flow, measured every hour (Alghamdi, 2019). There is paper that study about ARIMA and its variations for its application and compatibility with traffic predictions on various types of roads (Gavirangaswamy, 2013). Xu predicts road traffic state using ARIMA and Kalman filter that can achieve high accuracy prediction (Xu, 2017).

From the description above, the writer has an idea to conduct research using the ARIMA method on each number of cars and the number of motorcycles as outlined in his research entitled Forecasting the Number of Vehicles in Indonesia Using Auto Regressive Integrating Moving Average (ARIMA) Method.

2. Methodology

ARIMA model is a model that can be used to predict future values. The advantage of ARIMA is that it is flexible (follows data patterns) and has a high degree of forecasting accuracy in the short term. All data in the ARIMA model analysis are assumed to be stationary.

In ARIMA model divided into three groups, namely: the Autoregressive (AR) model, Moving Average (MA), and the mixed Autoregressive Moving Average (ARMA) model which has the characteristics of the first two models. In ARIMA (p, d, q), p is the order in the AR model, d is the integrated (differenced) order, that is differencing on non-stationary data, and q is the MA order (Montgomery, et al, 2008).

The AR (p) model is denoted as follows,

\[ Y_t = \delta + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + \epsilon_t \]  

where \( Y_t \) is the predicted variable or the dependent variable. \( \delta \) is a constant. \( Y_{t-1}, Y_{t-2}, Y_{t-3}, \ldots, Y_{t-p} \) are independent variables which are lags of dependent variables. \( \phi_1, \phi_2, \phi_3, \ldots, \phi_p \) are autoregressive parameters. \( \epsilon_t \) is an error or error value that cannot be explained by the model.

The MA (q) model is denoted as follows,

\[ Y_t = \mu + \epsilon_t - \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2} - \cdots - \theta_q \epsilon_{t-q} \]  

where \( Y_t \) is the predicted variable or the dependent variable. \( \mu \) is a constant. \( \theta_1, \theta_2, \theta_3, \ldots, \theta_p \) are the moving average parameters. \( \epsilon_t \) is an error value that cannot be explained by the model. \( \epsilon_{t-1}, \epsilon_{t-2}, \epsilon_{t-3}, \ldots, \epsilon_{t-q} \) are the error values at time \( t \).

The ARMA model (p, q) which is a mixture of the AR and MA models is denoted as follows,

\[ Y_t = \delta + \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} + \epsilon_t - \theta_1 \epsilon_{t-1} - \phi_q \epsilon_{t-q} \]  

ARIMA model which is the application of the AR, MA, and ARMA models shows that the data already have stationary properties. But in reality there are often data that are not stationary, to achieve stationarity in the data it is necessary to do a differencing process on the data. If the data goes through the differencing process \( d \) times to become stationary, then the data is said to be non-stationary homogeneous at level \( d \). This model with stationary data is called the ARIMA model. Thus, if the data has been stationary at differencing \( d \) times and will apply ARMA \((p, q)\), then the resulting model is ARIMA \((p, d, q)\).

### 3. Model Foundation

#### a. Stationery of Data

The number of cars and the number of motorcycles registered at the Central Statistics Agency in 2001-2019 with a total of 19 observations. Figure 1 presents the time series graph of the data. Based on Figure 1, it appears that the number of cars and motorcycles has increased from year to year. A sharp increase occurred in the number of motorbikes that increased almost 7 times from 2001 to 2019, while the increase in the number of cars did not seem too significant. Figure 2 presents the results of trend analysis of cars and motorcycles.

![Fig. 1. Time series graph cars of and motorcycles](image)

Based on the plot and trend analysis of time series data in Figure 2, it appears that the data on the number of cars and the number of motorcycles tends to increase. Shows that the spread of data (means) is not constant. Therefore, the data is not stationary.

![Fig. 2. Trend graph of cars and motorcycles](image)

#### b. Model Identification

Based on Figure 3 and Figure 4 show the ACF and PACF plots that show non-stationary data because the value of the autocorrelation coefficient in time lag 1 is quite significant from 0 and decreases slowly while the coefficient of partial autocorrelation decreases dramatically and approaches 0 after time -lag 1. In addition, the autocorrelation and partial autocorrelation coefficients are outside...
the red line (Bartlett). This shows that the data on the number of cars and the number of motorcycles is not stationary. To overcome the problem of data unstation, a differentiation process is carried out. So we get the results in Figure 5.

From the trend plot, the first difference in Figure 5 shows that the data is stationary because the average number of cars and the number of motorcycles does not move freely in a certain period of time and has a small enough variance value and the actual value is approaching a linear line. After differencing, the ACF and PACF plots are displayed again in Figure 6 and Figure 7.

Based on Figure 6 and Figure 7 shows that the value of the autocorrelation coefficient and partial autocorrelation coefficient on the car decreases exponentially after time lag 1 while the autocorrelation coefficient and partial autocorrelation coefficient on the motor decreases rapidly after time-lag 2, the temporary model obtained is ARIMA (1,1,1) for cars and ARIMA (2,1,2) for motorcycles.
c. Parameter Estimation

The temporary model is tested for its parameters to get the best model with the results shown in table 1 and table 2:

**Table 1. Estimation results of ARIMA models number of cars**

| Model            | Significance Parameter | White Noise | Normality | MSE       |
|------------------|------------------------|-------------|-----------|-----------|
| ARIMA(1,1,1)     | x                      | √           | √         | 5.48×10^{10} |
| ARIMA(1,1,0)     | √                      | √           | √         | 5.72×10^{10} |
| ARIMA(0,1,1)     | √                      | x           | x         | 2.87×10^{11} |

**Table 2. Estimation results of ARIMA models number of motorcycles**

| Model            | Significance Parameter | White Noise | Normality | MSE       |
|------------------|------------------------|-------------|-----------|-----------|
| ARIMA(2,1,2)     | √                      | √           | √         | 2.51×10^{12} |
| ARIMA(2,1,1)     | x                      | √           | √         | 3.60×10^{12} |
| ARIMA(2,1,0)     | x                      | √           | √         | 3.61×10^{12} |
| ARIMA(1,1,2)     | x                      | √           | √         | 3.46×10^{12} |
| ARIMA(1,1,1)     | x                      | √           | √         | 3.46×10^{12} |
| ARIMA(1,1,0)     | √                      | √           | √         | 3.31×10^{12} |
| ARIMA(0,1,2)     | √                      | √           | √         | 1.26×10^{13} |
| ARIMA(0,1,1)     | √                      | √           | √         | 1.72×10^{13} |

Based on Table 1 shows that ARIMA (1,1,0) is the best model. While Table 2 shows that ARIMA
d. Forecasting

The final step is forecasting using a model that has been tested before. Forecasting used for the number of cars is the ARIMA model (1,1,0) and for the number of motorcycles is the ARIMA model (2,1,2). The results of forecasting the number of cars and the number of motorcycles until 2030 are as follows:

| Year | Forecast of Cars | 95% Limits | Year | Forecast of Motorcycles | 95% Limits |
|------|------------------|------------|------|-------------------------|------------|
|      | Lower            | Upper      |      | Lower                   | Upper      |
| 2020 | 18.026.465       | 18.057.539 | 2020 | 132.917.660             | 133.019.726|
| 2021 | 18.805.407       | 18.776.953 | 2021 | 139.098.942             | 143.910.029|
| 2022 | 19.575.293       | 21.313.517 | 2022 | 145.135.720             | 152.157.867|
| 2023 | 20.336.229       | 22.869.514 | 2023 | 151.031.298             | 161.105.996|
| 2024 | 21.088.318       | 24.503.403 | 2024 | 156.788.923             | 170.875.129|
| 2025 | 21.831.664       | 26.205.348 | 2025 | 162.411.778             | 181.479.691|
| 2026 | 22.566.368       | 27.967.751 | 2026 | 167.902.979             | 192.887.783|
| 2027 | 23.292.530       | 29.784.501 | 2027 | 173.265.581             | 205.049.023|
| 2028 | 24.010.251       | 31.650.545 | 2028 | 178.502.572             | 217.906.516|
| 2029 | 24.719.627       | 33.561.602 | 2029 | 183.616.876             | 231.402.212|
| 2030 | 25.420.757       | 35.513.990 | 2030 | 188.611.352             | 245.479.487|

Based on Table 3 shows that the number of cars and the number of motorcycles has continuously increased as in previous years. On the Table 5 and Table 6 shows the results of the MAPE, MAD, and MSD values. The model used to get the MAPE, MAD, and MSD values is the ARIMA model (1,1,0) for the number of cars and the ARIMA model (2,1,2) for the number of motorcycles.

| Statistics | Value          |
|------------|----------------|
| MAD        | 1.503.686      |
| MSE        | 2.265.092.297.823 |
| MAPE       | 7.01%          |

| Statistics | Value |
|------------|-------|
| MAD        | 11.471.877 |
| MSE        | 132.379.670.689.497 |
| MAPE       | 7.24% |

4. Conclusion

Based on the forecasting results with the ARIMA method, it can be concluded that:

a) The best ARIMA model for estimating the number of cars is ARIMA (1,1,0) with MAD 1.503.686, MSE 2.265.092.297.823 and MAPE 7.01%.

b) The best ARIMA model for estimating the number of motorcycles is ARIMA (2,1,2) with an MAD 11.471.877, MSE 132.379.670.689.497 and MAPE 7.24%.

c) Forecasting results on the number of cars and motorcycles shows that the number of the vehicles is continue to increase from year to year

d) The number of cars and the number of motorcycles forecasting models have a small MAPE value, which is less than 10%. This means the models have high accuracy.

There are many factors that influence the number of cars and motorcycles. Fuel prices, gross
domestic product, population, road growth are important factors. Therefore, forecasting the number of cars and the number of motorcycles can not only use the time series model but with a causal model.

References
[1] Alghamdi, Taghreed & Elgazzar, Khalid & Bayoumi, Magdi & Sharaf, Taysseer & Shah, Sumit.: Forecasting Traffic Congestion Using ARIMA Modelling. International Wireless Communications and Mobile Computing Conference (2019)
[2] Farizal, Alrasyid, H., and Rahman, A: Model peramalan konsumsi premium Indonesia dengan regresi linear berganda. Jurnal Ilmiah Teknik Industri UMS. Vol 13, No. 02 (2014)
[3] Gavirangaswamy, Vinay B & Gupta, Gagan & Gupta, Ajay & Agrawal, Rajeev.: Assessment of ARIMA-based Prediction Techniques for Road-Traffic Volumes. International Conference on Management of Emergent Digital EcoSystems (2013)
[4] Montgomery, Douglas & Jennings, Cheryl & Kulahci, Murat. Introduction to Time Series Analysis and Forecasting. John Wiley & Sons, Inc. Canada (2008)
[5] Pangestu, F.: Prediksi Jumlah Kendaraan Bermotor di Indonesia Menggunakan Metode Average-Based Fuzzy Time Series Models. Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer. Vol. 2, No. 9. (2018)
[6] Rizki, Febriasto.: Peramalan Jumlah Mobil Pribadi yang Berada di Kota Surabaya. Institut Teknologi Sepuluh Nopember. Surabaya. Indonesia (2017)
[7] Salim, Agus.: Teori dan Paradigma Penelitian Sosial. Tiara Wacana, Yogyakarta. Indonesia (2006)
[8] Shakti, Sana & Hassan, Mohan & Zhenning, Yang & Caytiles, Ronnie & Iyenger, N Ch. S. N.: Annual Automobiles Sales Prediction Using ARIMA Model. International Journal of Hybrid Information Technology. Vol. 10, No. 6 (2017)
[9] Sudjana.: Metoda Statistika. Tarsito, Bandung. Indonesia (2005)
[10] Vaghasia, Sukruti.: An Approach of Traffic Flow Prediction Using ARIMA Model with Fuzzy Wavelet Transform. University of Windsor. Canada (2018)
[11] Xu, Dong-wei & Wang, Yong-dong & Jia, Li-min & Qin, Yong & Dong, Hong-hui.: Real-time Road Traffic State Prediction Based on ARIMA and Kalman Filter. Frontiers of Information Technology & Electronic Engineering (2017)