Presented at the DLSU Research Congress 2016
De La Salle University, Manila, Philippines
March 7-9, 2016

Modelling and Forecasting Light Rail Transit Line 1 Patronage Using Seasonal Autoregressive Integrated Moving Average Method

Mary Joyce Bernadette B. Asutilla[1,*], Monica Sanhanda[1] and Rechel G. Arcilla[1]
[1] *Mathematics Department, De La Salle University*
**Corresponding Author: iva.asutilla@gmail.com*

Abstract: Public mass transportation is an important sector in a country for it provides accessible, efficient yet affordable means of transportation for the people. Various means of transportation are available for the public and in the National Capital Region of the Philippines, the Light Rail Transit Line 1 (LRT-1) is one of those means of transportation. It runs along Taft Avenue to North EDSA, servicing thousands of passengers daily. For its railway operations and planning, the Light Rail Transit Authority (LRTA) uses the Annual Average Growth Rate method in forecasting LRT-1 patronage. The method is inappropriate for forecasting since seasonality was observed for LRT-1 patronage from January 2000 to August 2015. This study considered the Seasonal Autoregressive Integrated Moving Average (SARIMA) method in modelling and forecasting LRT-1 patronage to address the presence of seasonality in the data. Measures of the forecast accuracies, namely the mean error (ME), mean absolute deviation (MAD), mean percentage error (MPE) and mean absolute percentage error (MAPE) showed that the proposed *ARIMA* $(0,1,1)\times(0,1,1)_{s=12}$ with $\mu=0$ model is a good model to forecast LRT-1 patronage for September 2015 to August 2016 using in-sample evaluation. Hence, the proposed SARIMA model and the forecasts based from the model is recommended to help LRTA in its operations, considering that LRT-1 patronage peaks during March, September until December due to the enrollment period and Christmas season respectively. The proposed SARIMA model suggests that maintenance for LRT-1 may be done during April since the demand for its services is low as compared to other months. Further, the model is recommended to further help LRTA in its operations to provide better services to the public.

Key Words: LRT-1 patronage; forecasting; SARIMA method

1. INTRODUCTION

The Light Rail Transit Line 1 (LRT-1) is one of the three main railway systems in Metro Manila. It is the oldest railway system which is 19 km long along Taft Avenue, Rizal Avenue and North EDSA. It is governed by the Light Rail Transit Authority (LRTA), a government-owned and controlled corporation under the authority of the Department of Transportation and Communications (DOTC). As its vision and mission, LRTA aims to cater to the needs of the public for a rapid transit system aimed at providing world class light railway systems and faster basic mobility service to the general public. Through the

Accomplishment Report presented by LRTA every year, the public is informed of the accomplishments of the agency. As of 2014, several performance indicators were met for LRT-1 such as the expected increase in its patronage and availability of trains. However, the capacity utilization and load comfort ability of a train were not met since its annual average load factor for the said year was 98.01% which is not within the range of 90 to 95%. There was a growth of 3.71% in its patronage for 2014 and the insufficient number of trains to accommodate passengers has resulted in passenger traffic congestion.

The last study that dealt with the railway transportation in Metro Manila was in 2009 when Pineda and Sarmiento forecasted the patronage for the Metro Rail Transit (MRT-3) using the Univariate Box-Jenkins method. Results of their study have shown that the monthly MRT-3 patronage for 2001 to 2008 had an increasing trend and the final model proposed may be used to validate the policies and projects of MRT-3.

With the emerging need for research in the field of public mass transportation, the study intended to provide the best SARIMA model to provide forecasts for the LRT-1 patronage for September 2015 to August 2016. The study aimed to contribute to the body of knowledge in the transportation sector by proposing a forecasting method which may help LRTA in their railway operations and their decision-making process.

2. METHODOLOGY

2.1. Data

LRT-1 patronage from January 2000 to August 2015 was used which is the aggregated number of successful entries collected from the sensors in each automatic gate of each station. The counts were sent every 15 minutes to the station servers which were then aggregated into a day, then, into a month.

2.2. SARIMA Modelling Procedure

Two preliminary tests were conducted to determine whether seasonality is present in the data. The X11 procedure in SAS 9.3 conducted two *F* tests to detect stable and moving seasonality while a Kruskal-Wallis test was conducted to detect for the presence of identifiable seasonality. Once seasonality was detected, the SARIMA model was considered.

SARIMA is a generalized ARIMA model that considers the presence of seasonality where the model is given by

$$\Phi_p(B^s)\varphi_p(B)(1-B^s)^D(1-B)^dY_t = \mu + \Theta_Q(B^s)\theta_q(B)\varepsilon_t \quad \text{(Eq. 1)}$$

where:

$Y_t \sim ARIMA(p, d, q) \times (P, D, Q)s$

p = order of ordinary autoregressive
d = order of ordinary differencing
q = order of ordinary moving average
P = order of seasonal autoregressive
D = order of seasonal differencing
Q = order of seasonal moving average
s = length of seasonality

The X12 procedure in SAS 9.3 was used to automatically generate the best SARIMA model for the data. Residuals of the proposed best SARIMA model were checked whether the assumption for white noise was satisfied. Forecasts generated by the best SARIMA model were visually presented.

Measures of the forecast accuracies of the model were considered to determine if the model is a good forecasting model. Values that are near to 0 for the ME, MAD, and MPE while a MAPE value of at most 10% are preferred.

All statistical analyses performed used a 5% level of significance.

3. RESULTS AND DISCUSSION

Despite internal and external factors affecting the choice of the public, LRT-1 patronage has been continuously increasing (Figure 1). The monthly LRT-1 patronage from January 2000 to August 2015 showed an average of 10.98 million passengers where the lowest patronage of 4.25 million was reported last August 2000. This sudden decrease in LRT-1 patronage was due to the strike that the METRO INC. employees staged last July 25 to August 2 of 2000 that paralyzed the operations of LRT-1. In addition, the consequences of the strike affected the efficiency of the services provided by LRT-1 and the 16-year Operating and Maintenance Agreement with METRO INC. was not renewed.

ISSN 2449-3309

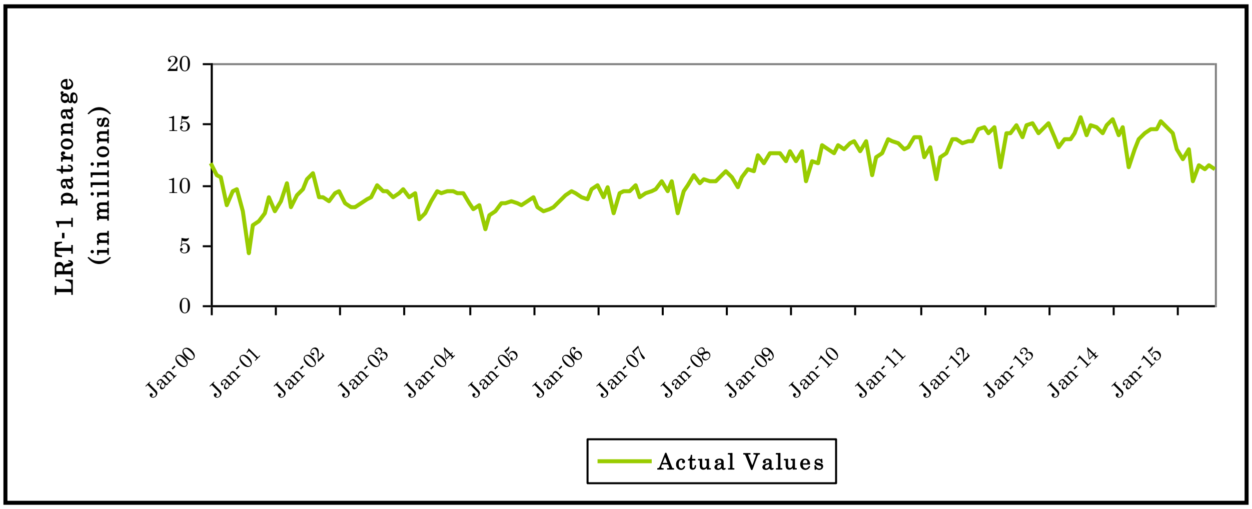

Fig. 1. LRT-1 patronage in millions for January 2000 to August 2015

Seasonality is an important component in a time series data since its presence signifies that the data has predictable changes that occur every year. Formal tests for the presence of identifiable seasonality were conducted. With the Kruskal-Wallis test, it was verified that in the monthly LRT-1 patronage for January 2000 to August 2015, identifiable seasonality was present (see Table 1). The tests confirmed the observed seasonality since LRT-1 patronage peaks during the Christmas season in the months of September to December since Divisoria and Baclaran, popular areas where affordable goods are available, are accessible using this mode of transportation. In addition, LRT-1 patronage increases during March as this is the usual enrollment period for various universities.

Table 1. Test for seasonality

Tests	df	Value
Stable Seasonality F-test	(11,176)[1]	15.78*
Moving Seasonality F-test	(14,154)[2]	2.08*
Kruskal-Wallis Chi-square test for presence of identifiable seasonality	11	96.40*

[1]df of (between months, error) [2]df of (between years, error)
*significant at 0.05 level

Since the presence of seasonality was verified, the X12 procedure was conducted to generate the best SARIMA model for the LRT-1 patronage. The Schwarz Bayesian criterion (SBC) statistic was used to determine which among the recommended ARIMA models would be the best for the LRT-1 patronage. The model with the lowest SBC statistic is preferred and the proposed *ARIMA* $(0,1,1)\times(1,0,1)_{s=12}$ had the lowest SBC statistic (see Table 2). The recommended SARIMA model was compared with the default *ARIMA* $(0,1,1)\times(0,1,1)_{s=12}$ model to determine which is the final best SARIMA model for the time series data. Comparison between the two SARIMA models was based on the presence of unit roots, non-seasonal over differencing, and insignificant parameters where absence of the three criteria was preferred. It was concluded that the *ARIMA* $(0,1,1)\times(0,1,1)_{s=12}$ model is the best model since its SBC value of 426.7753 is lower as compared to the SBC value of 455.8342 for the *ARIMA* $(0,1,1)\times(1,0,1)_{s=12}$ model.

Table 2. Best five ARIMA models

Rank	Estimated Model						SBC
1	(0,	1,	1)	(1,	0,	1)	455.8342
2	(1,	1,	0)	(1,	0,	1)	458.1400
3	(1,	1,	1)	(1,	0,	1)	461.0249
4	(0,	1,	2)	(1,	0,	1)	461.0364
5	(2,	1,	0)	(1,	0,	1)	461.4418

Parameter estimates for the proposed *ARIMA* $(0,1,1)\times(0,1,1)_{s=12}$ model indicated that the constant term μ is not significantly different from 0 (p-value = 0.7873). Thus, another model *ARIMA* $(0,1,1)\times(0,1,1)_{s=12}$ with $\mu=0$ was considered. Table 3 presents the parameter estimates of the best model which is similar to the final SARIMA model proposed by Pineda and Sarmiento (2009) for the MRT-3 patronage. The LRT-1 patronage, however, did not need any further data transformation as compared to the MRT-3 patronage where the quartic transformed data was used.

Table 3. Parameter estimates for the SARIMA model

Parameter	Estimate	Standard Error	t Value
θ_1	0.3604	0.0712	5.06*
Θ_1	0.5159	0.0672	7.68*

*significant at 0.05 level

The residuals of the SARIMA model were evaluated to determine whether these are not significantly different from white noise. Clearly, the assumption that the residuals are not significantly different from white noise was satisfied ($X^2_{(46)}$ = 34.16, p-value = 0.9013). Hence, the final SARIMA model for the LRT-1 passenger ridership is expressed as

$$(1-B)(1-B^{12})\,Y_t = (1-0.3604B)(1-0.5159B^{12})\varepsilon_t \quad \text{(Eq. 2)}$$

Measures of forecast accuracies for the chosen SARIMA model showed a MAPE of 5.6328% while the values for its ME, MAD, and MPE were close to 0, indicating the model provided accurate forecasts (Table 4).

Table 4. Measures of forecast accuracies

Statistic	Value
ME	0.0489
MAD	0.5894
MPE	0.0039
MAPE	5.6328%

Further, forecasts for LRT-1 patronage for September 2015 to August 2016 are presented in Table 5 and Figure 2 where the forecasts were within a narrow 95% confidence interval, implying the precision of the forecasted values.

Table 5. Forecasts for LRT-1 patronage in millions

Month and Year	*ARIMA* $(0,1,1)\times(0,1,1)_{s=12}$
September 2015	11.56
October 2015	11.91
November 2015	11.43
December 2015	11.50
January 2016	11.09
February 2016	10.07
March 2016	10.58
April 2016	8.25
May 2016	9.52
June 2016	9.69
July 2016	10.23
August 2016	9.87

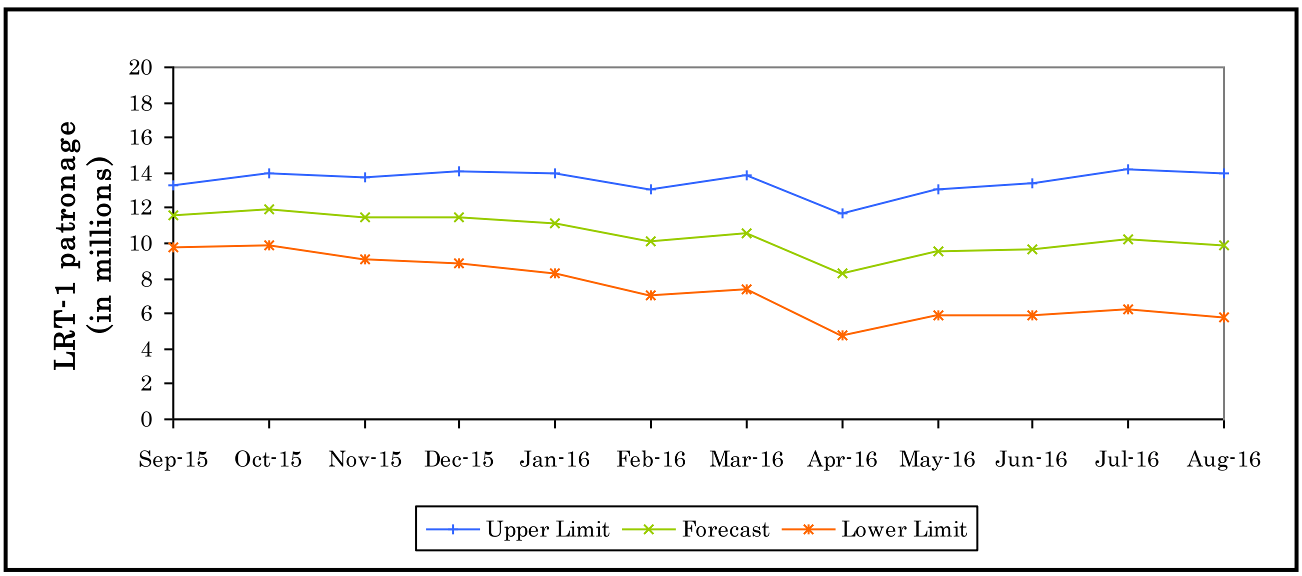

Fig. 2. Forecasted LRT-1 patronage and 95% confidence interval in millions for September 2015 to August 2016

ISSN 2449-3309

The seasonal factors for January 2011 to August 2015 are presented in Figure 3 and the pattern has confirmed that the lowest demand for LRT-1 is during the month of April. This behavior is expected since most of the universities are on summer break. Consequently, the students who usually ride the light rail vehicles are not accounted for during this month. It is also evident that the lowest forecasted value was reported during April 2016. This behavior can be explained by the circumstance that the summer season is during this month. Most of the universities that are accessible by this means of transportation do not have their regular class schedules. Hence, LRTA may conduct the various maintenance procedures for its trains and other facilities during the month of April.

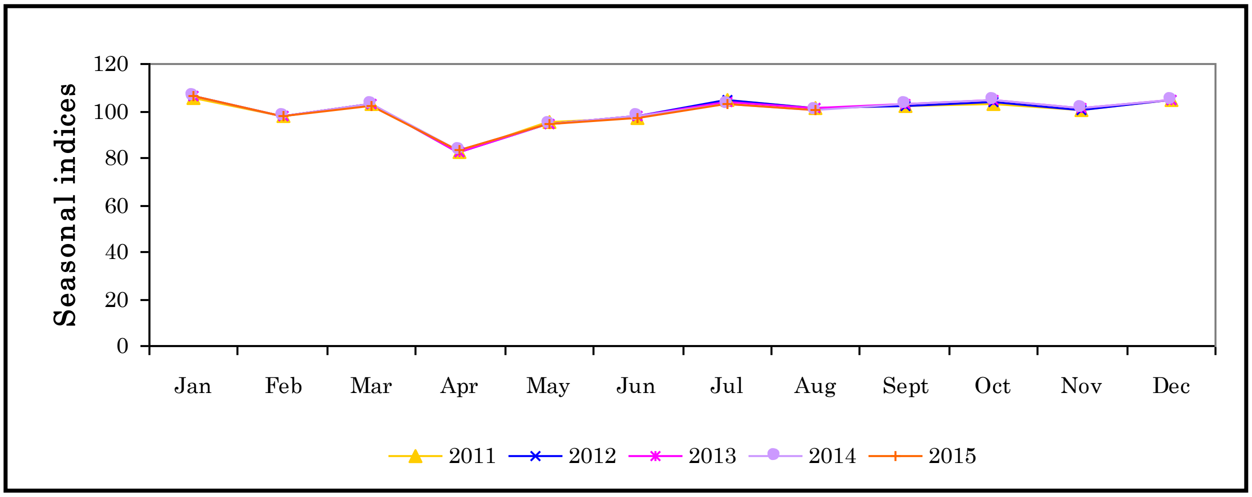

Fig. 3. Seasonal indices from January 2011 to August 2015

4. CONCLUSIONS

The aggregated number of successful LRT-1 passenger entry across all stations from January 2000 to August 2015 showed that seasonality was identifiable and present as confirmed by the Kruskal-Wallis test for seasonality. This behavior of the LRT-1 patronage is not accounted for by the current policy and procedure in estimating future patronage conducted by the LRTA which is important in the decision-making process of the agency. It was also observed that the LRT-1 patronage peaks on March and from September to December due to the enrollment period and Christmas season, respectively. Maintenance of the services may be done during April as the demand for the services offered by LRT-1 is not as high as compared to the other months. An increasing trend was observed, implying that there is a need for LRTA to further increase the capacity of their trains and other services facilities.

The final *ARIMA* $(0,1,1)\times(0,1,1)_{s=12}$ with $\mu=0$ model provided a more parsimonious model that would give better forecasts with MAPE of 5.6328% that is less than 10% and values of 0.0489, 0.5894, and 0.0039 for the ME, MAD, and MPE, respectively, that are close to 0. The model considered accounts for the seasonality in the data which is an essential behavior of the LRT-1 patronage which LRTA needs to address.

It is recommended for future studies to explore on the robustness of the model by considering training and testing sets. Other railway patronage may also be considered in forecasting and modelling using the SARIMA method to determine if the proposed *ARIMA* $(0,1,1)\times(0,1,1)_{s=12}$ model suits the railway patronage for the different railway systems. It is also recommended for future studies to consider whether the number of functional automatic gates, number of functional trains operating monthly, and the location of a train station have an impact on its patronage.

5. ACKNOWLEDGMENTS

The authors would like to thank the Open Data Policy of DOTC that made available the monthly LRT-1 patronage data. The authors would also like to recognize the assistance extended by Mrs. Vicky Pacis-Apacible, Officer-in-Charge of the Public Information Division of DOTC and Ms. Remer Cuarez, Public Relations Officer I of LRTA regarding clarifications on the data used.

6. REFERENCES

Light Rail Transit Authority. (2014). Year-end accomplishment report [Electronic version]. Retrieved December 18, 2015, from http://www.lrta.gov.ph/images/upload/2014%20Year-End%20Accomplishment%20Report.pdf

Lrta.gov.ph. (n.d.). Company Profile. Retrieved December 18, 2015, from the Light Rail Transit Authority Website http://www.lrta.gov.ph/index.php/2014-05-21-01-05-51/2014-05-22-02-12-18/2014-05-22-02-13-21

Pineda, S., & Sarmiento, A. (2010, June). Forecasting Metro Rail Transit 3 ridership using univariate Box-Jenkins methodology. Paper presented at the 30th International Symposium on Forecasting, San Diego, CA. Retrieved December 18, 2015, from Https://forecasters.org/submissions10/PINEDASTEPHANIEISF2010.pdf

ISSN 2449-3309