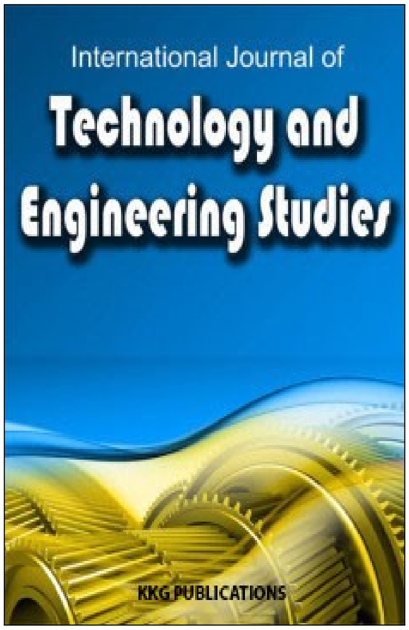

Key Knowledge Generation

Publication details, including instructions for author and subscription information:
http://kkgpublications.com/technology/

Forecasting the Air Passenger Volume in Singapore: An Evaluation of Time-Series Models

GUO RUI [1], ZHONG ZHAOWEI [2]

[1, 2] School of Mechanical and Aerospace Engineering,
Nanyang Technological University, Singapore

Published online: 22 June 2017

To cite this article: G. Rui and Z. Zhaowei, "Forecasting the air passenger volume in singapore: An evaluation of time-series models," *International Journal of Technology and Engineering Studies,* vol. 3, no. 3, pp. 117-123, 2017.
DOI: https://dx.doi.org/10.20469/ijtes.3.40004-3

To link to this article: http://kkgpublications.com/wp-content/uploads/2017/3/IJTES-40004-3.pdf

KKG PUBLICATIONS

International Journal of Technology and Engineering Studies
vol. 3, no. 3, pp. 117-123, 2017

IJTES

FORECASTING THE AIR PASSENGER VOLUME IN SINGAPORE: AN EVALUATION OF TIME-SERIES MODELS

GUO RUI [1*], ZHONG ZHAOWEI [2]

[1, 2] School of Mechanical and Aerospace Engineering, Nanyang Technological University, Singapore

Keywords:
Air Passenger Volume
Time-Series Models
Long-Term Forecasting

Received: 02 January 2017
Accepted: 15 March 2017
Published: 22 June 2017

Abstract. This paper explores various methods to predict air passenger movements and analyzes and compares the relative results of corresponding models. Due to the increasing development of air transport technology, air passenger movements have been growing dynamically. Therefore it is necessary to have a good forecasting model suitable for Singapores situation. 8 time-series models were simulated for 18 years of prediction from 1998 to 2015 in the study and were compared based on their forecasting error measurements. Finally, appropriate models for Singapores situation are recommended. Afterward, forecasting for the next 18 years is conducted to have an idea about the future development. Accurate forecasting information will lead to appropriate timing for facility construction with minimum effect on service. In addition, the long-term forecasting will also provide information for aircraft ordering and design in consideration of bigger aircraft to carry more passengers

INTRODUCTION

Background

Along with Singapore's rapid development, the country has become a financial center in Asia. Concurrently, Singapore has been attracting more and more tourists due to the good image. Singapore Changi airport has been established into one of the largest transportation air hubs in Southeast Asia [1]. It has been reported that Changi airport has been ranked the world's best airport for 4 years [1].

Moreover, it has been estimated that air transport will be dominant in this century for both passengers and freight [2]. Hence, to uphold the service standard, one important contribution is to expand the facilities promptly to prevent congestion and delay and meet customer satisfaction. In order to carry out such projects to maximize profit and minimize loss, accurate forecasts are necessary and required at a strategic level.

On the other hand, air traffic transport is also closely related to economic growth. Therefore accurate forecasting can also provide practical information for society's economic development [3].

LITERATURE REVIEW

Several time-series approaches have been studied ranging from the Holt's method, linear regression model, non-linear regression model and ARIMA model. Bermudez and his peers have used Holt-Winters method to forecast UK air passenger volume [4]. Profillidis has made attempt to forecast the demand of Rhodes airport by using a linear regression model and a polynomial trend model, showing a good match [5]. Rodrigo employed Holt-Winter model to conduct forecasting for air passengers volume of Sao Paulo International airport [6]. Emily and his peers have conducted forecasting of Chinese tourist arrivals in Australia by ARIMA model [7]. Wai and his peers forecasted Hong Kong airportls passenger volume by using SARIMA and ARIMAX models, which considers seasonality [8]. According to Howard and his colleagues, univariate forecasting outperforms multivariate econometric modelling for long-term forecasting [9]. Andreoni and his colleagues have utilized ARIMA model for air traffic forecasting [10].

Objectives

Since the long-term planning for Singapore air passenger volume can assist on enhancing aviation industry development, it is useful to find the best performed univariate model for Singapore.

In addition, although linear trend and ARIMA models have been explored in depth, other time-series methods have not been studied sufficiently. Moreover, all the time-series models have not been evaluated thoroughly. Under such circumstances, this paper explored more time-series modelling and compared the performance of the models for Singapore's situation for the first time. Moreover, in this study four error measurements are employed together to evaluate the performance of different models for the first time. By comparison, the most proper model was chosen for air traffic forecasts. In consideration

*Corresponding author: Guo Rui
†Email: GUORUI@ntu.edu.sg

of large contribution of Singapore's air traffic to economic growth, because the air passenger volume can reflect the status of Singapore air traffic development, this paper aims to forecast air passenger growth for next 18 years. The forecasting results, then, can be used to deduct the appropriate timing for infrastructure expansion.

METHOD AND MATERIALS

Data Collection

To carry out comparison of performance of different forecasting models, the numbers of yearly passenger movements in Changi airport from 1998 to 2015 were collected from Wikipedia and were used to build the models, as shown in Figure 1 [11]. It is shown that the passenger volume was growing overall, but in specific years the number decreased.

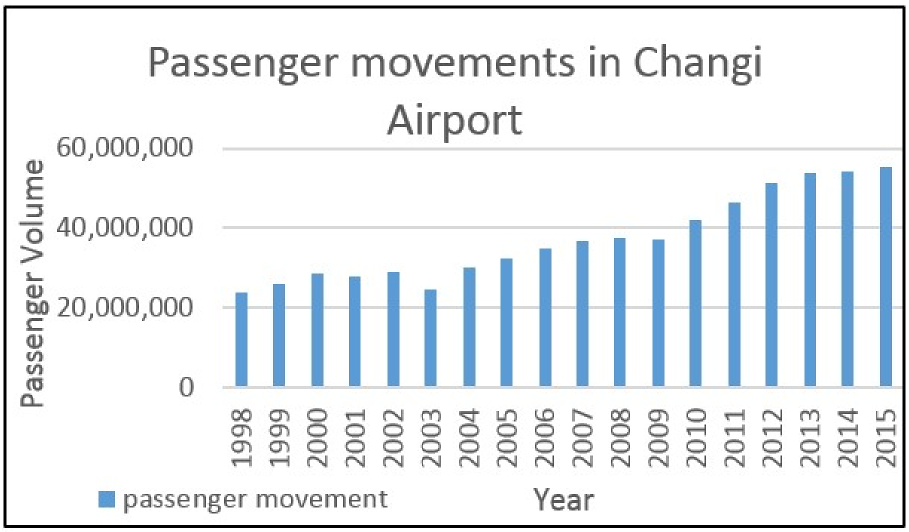

Fig. 1. Actual air passenger movements from 1998 to 2015 [2]

Time Series Analysis

The Holt's Method

Holt's method is defined as exponential smoothing of exponential smoothing [12] [13]. The general equations are demonstrated in the following manners [12]:

$$F_{(t+1)} = I_t + S_t \tag{1}$$

$$I_t = +(1-\alpha)(I_{t-1} + S_{t-1}) \tag{2}$$

$$S_t = \beta(I_t - I_{t-1}) + (1-)S_{t-1} \tag{3}$$

I_t is expected level of the time series while S_t is the expected rate of increase or decrease per period [5]. Equation (1) shows that the forecasting value is the addition of expected level I_t and expected rate S_t. To calculate the value of I_t andS_t by using Equation (2) and Equation (3) respectively, the two smoothing constants α and β need to be predefined. The values of α and β were set arbitrarily, the Mean Absolute Deviation (MAD) and Mean Squared Error (MSE) were minimized in order to optimize α and β [14], [15]. Consequently, α was modified to 0.64 and β was 0.28.

Linear Trend Model

The general form isy_t= $b_0 + b_1 t_i$, where t_i represents the time. The linear trend graph was plotted by following the actual passenger volume in Excel. The equation for forecasting this data set was estimated to be:

$$y_t = -3834403386.8 + 1929614.608 t_i \tag{4}$$

The coefficient of determination R^2 was 0.92, showing a relatively good fit of the actual movement preliminarily. The Mean Absolute Percentage Error (MAPE) was 7.29%. According to Lewis, when the MAPE value is less than 10% for a forecasting model, the forecasting performance is deemed as highly accurate [16].

The t values of the coefficient and intercept are much larger than the critical t-value, 3.965, at 99.5% confidence level for 17 degrees of freedom. Moreover, p-values are much smaller than 0.05, which indicates rejection of null hypothesis. Thus, the linear model showed in Equation (4) is statistically significant.

Quadratic Trend Model

The general form is $y_t = b_0 + b_1 t_i + b_2 t_i^2$. With trend addition, the quadratic trend equation was estimated to be:

$$y_t = 91963.66 t_i^2 - 3.6710^8 t_i + 3.6610^{11} \quad (5)$$

R^2 was 0.967, which implies a highly fit regression model. Figure 2 illustrates quite a good match. Especially some points show that the estimation errors are close to 0 (e.g. demand in year 2005 and 2014). According to Table 1, the t value is observed to be relatively larger than the critical t-value, 3.965, at the 95% confidence level for two tails. The *p*-value is also noticed to be less than 0.01, which shows statistical significance difference from the null hypothesis (the data represent only random fluctuations around the sample mean).

TABLE 1
COEFFICIENT DETAILS OF QUADRATIC TREND MODEL

	St-Error	t-Stat	*p*-value
Intercept	8.2210^{10}	4.459	0.000459
Year	81,892,638	-4.483	0.000438
Year-square	20,407	4.507	0.000418

Cubic Trend Model

The general form is $y_t = b_0 + b_1 X_i + b_2 X_i^2 + b_3 X_i^3$. Following the same procedure as quadratic model, the best fitted equation with R^2 of 0.968 is found to be:

$$y_t = -3099.416x^3 + 18694722.38x^2 - 37693406812x + 2.5331310^{13} \quad (6)$$

Exponential Trend Model

The general model is$y_t = \alpha e^{\beta x} or ln y_t = ln\alpha + \beta x$. It was discovered that the equation best functioned for this particular dataset is:

$$ln y_t = -85.345 + 0.0512 t_i \quad (7)$$

R^2 was 0.941. The *p* value of Equation (7) is less than 0.01 to a great extent, indicating that the null hypothesis is rejected. The t value also shows the model is statistically significant.

Logistic Growth Model

The general form of logistic model is $Y_i = \frac{\beta_1}{1+exp(\beta_2+\beta_3 x_i)} + \epsilon_i$, where y_i represents the estimated passenger number at time x_i, β_1 is the asymptote towards which the passenger volume grows, β_2 reflects the initial passenger volume at initial time, and β_3 controls the growth rate of the passenger volume [17]. In our model, the assumption of error ϵ_i = 0 is made. This model was optimized using Gauss-Netwon method. Nonlinear regression was selected and the passenger volume was chosen for the response. After calculation, the equation was finalized as:

$$y_t = 4.4881810^9/(1 + exp(5.35252 - 0.0597719 * (t_i - 1998))) \quad (8)$$

R^2 of Equation (8) was 0.947.

ARIMA Model

The general form of an ARIMA (p, d, q) model is [12]:

$$\phi(B)(1-B)^d y_t = (B)\epsilon_t \quad (9)$$

where the ϕ_i is autoregressive parameter, the θ_i is moving average parameter and the ϵ' s are white noise error terms [3] [11]. In the ARIMA model, backward shift operator B $(B^m y_t = y_{t-m})$ is used. In Equation (10) ϕ (B) (ϕ (B) =1 $-$ $\emptyset_1 B - \emptyset_2 B^2 ... - \emptyset_P B^P$) represents the autoregressive part polynomial of order p and θ (B) $(\theta(B) = 1 - \theta_1 B - \theta_2 B^2 ... - \theta_q B^q)$ is the moving average part with highest order q. The parameter d indicates to which extent differencing is needed to transform a non-stationary model to stationary [12] [13].

The Iterative Box-Jenkins method was suggested to be utilized [18]. There are three steps inclusive of model identification, estimation and validation. In the model identification step, the data need to be verified whether it is stationary [18]. The time series y_t is (weakly or covariance) stationary if the mean and the variance of y_t are independent of time [12]. Therefore any impact on y_t will not have a permanent influence on the future development of the series [15].

The graphical judgement of Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots was employed. It has been found out that if the ACF plot expresses decay tendency or has an unusual large value, the data are of non-stationarity and need further improvement [18].

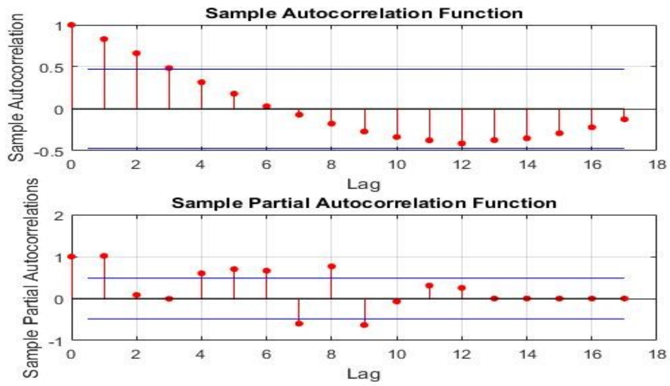

Fig. 2. ACF and PACF

Figure 2 shows the gradual decline of ACF values and outlier of the first PACF, which indicates non-stationary situation [19]. With the stationarity of the data achieved by adding two differencing terms, p and q need to be decided subsequently. The model ARIMA (3, 2, 2) is yield and simulated.

RESULTS ANALYSIS

Comparison of the Linear Model, Quadratic Model and Exponential Model

According to Profillidis, there are four criteria to evaluate the performance of forecasting model [13]. The linear trend model and the polynomial trend model are compared and assessed in this subsection. The term degree of divergence ($D_i = \frac{y_i - f_i}{y_i}\%$) is introduced by Profillidis [20], where f_i is the forecasted value of passenger demands for year i. $D_{mean} = \frac{1}{N}\sum_{i=1}^{N}(|D_i|)$ is used to evaluate the forecasting model [20]. In addition, the maximum degree of divergence max D_i(i=1, 2...N) [20], where N is the number of years accounted, is also one of the gauges. In this study, N= 18.

It was calculated that D_{mean}=7.54% and max D_i =13.25 % for the linear trend model, D_{mean}=4.26% and max D_i=8.32% for the quadratic model and D_{mean}=5.49% and max D_i=-26.08% for exponential model. The mean divergence for all three models is well below 10%, and it can be observed that the quadratic model outperforms the other two.

The mean divergence balance D_{bal} is another criterion for measurement. $D_{bal} = \frac{1}{N}|\sum_{i=1}^{N} D_i|$ can show the approximation of differences between the actual and estimated values. D_{bal}=0.37% is for the linear model while 0.33% for the polynomial model. Figure 3 illustrates that the forecasting values from the quadratic model are nearer to the real values as compared to the linear and exponential models. Profillidis also highlighted and conducted a physical comparison [20]. For the linear equation, the second-derivative equals to zero: $\frac{d^2y(t)}{dt} = 0$, which shows that no external impact is taken into consideration during this forecasting process. However, for the quadratic model, its second derivative is $\frac{d^2y(t)}{dt}$ = 91964, which indicates that an external impact exists with amplification effects on the model. The amplification effects can be due to GDP growths, more global cooperation and even increasing leisure tourists worldwide [20].

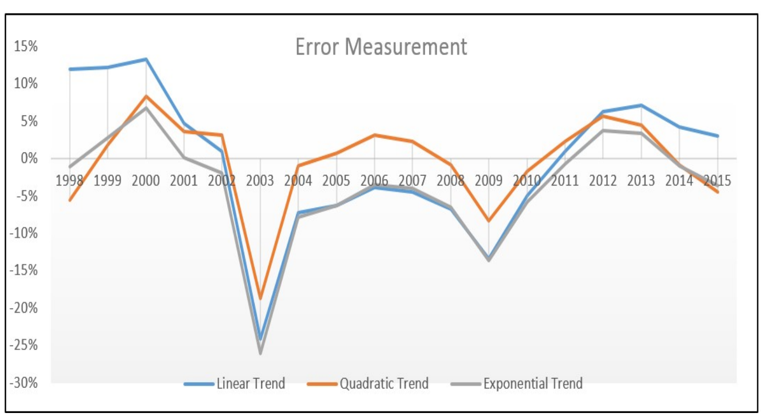

Fig. 3. Degree of divergence of linear and quadratic second-degree calibration

The R^2 of exponential trend model is lower than that of the linear and quadratic trend models. Furthermore, $\frac{Actual F}{Critical F}$ is 57 for the exponential trend model and 59.5 for the quadratic trend model. Since the ratio for the quadratic model is higher, the model is statistically better for this particular Singapore's situation.

Cubic Trend Model Analysis

The p-value of the cubic trend model is around 0.5, which is comparatively larger than 0.05. Moreover, the t-values for the three coefficients are all smaller than the critical t value, 0.689, at the 50% level. Hence, the null hypothesis (b_3=0) cannot be rejected, which indicates that the cubic trend model is not suitable for forecasting Singapore air passenger demands in the current situation. The reason may be that there is an explicit increasing trend of air passenger demands, which means the air passenger volume is growing at a rather stable rate in the sample data provided.

Logistic Growth Model

Due to large p-value, the model is not statistically significant. Hence, the model cannot illustrate the relationship of air passenger demands and time, and does not fit with this particular dataset. It proves from another perspective that the passenger demand is still on fast-growth trend. Hence, the logistic growth model is suggested to be applied when the market becomes mature or there is a capacity constraint.

ARIMA Model

The model is described by Equation (10): R^2 was 0.974. The t value was 12.239, which exceeds the critical value at the 5% critical level. Thus, the model was statistically significant.

$$y_t = 0.261 y_{t-1} - 0.205 y_{t-2} - 0.652 y_{t-3} - 1.479 e_1 + e_2 \quad (10)$$

Overall Comparison

Table 2 summarizes the error measurements for different models. The MAPE of the ARIMA (3, 2, 2) model is the smallest and the MAPE of the linear trend model is the largest. In addition, RMSE and D_{bal} are also the smallest. Thus, the ARIMA (3, 2, 2) model is considered the best performed model in time series models for Singapore air passenger forecasting. On the other hand, the RMSE, D_{bal} and the largest degree of divergence of Holt's method are the largest. Hence, Holt's method is not suitable for Singapore long-term passenger forecasts. On the other hand, the average growth rate of passenger demands is calculated. The average growth rate of the ARIMAR model and Holt's method is closest to the actual value 5.4%. Hence, during forecasting, it is essential to reflect the actual data's trend in the model. Figure 4 shows the comparison of results among time-series models. In general, the exponential model overestimated the passenger demands, which indicates Singapore air passenger hasn't reached the level of exponential growth rate. The linear model shows gradually slower growth rate as compared to the actual one. The ARIMA model, however, follows the actual trend tightly.

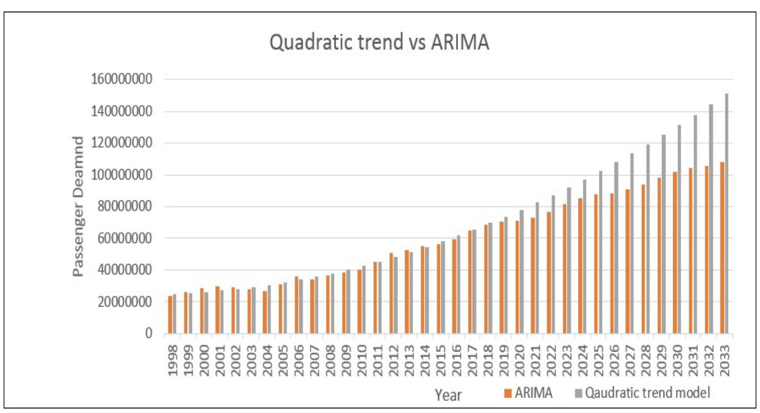

Fig. 4. Comparison of forecasting results among different models

Besides ARIMA model, the quadratic trend model is the best performed model. The value of largest degree of divergence of the quadratic model is smallest. This shows that the quadratic trend model is reliable in general. The result may indicate that the quadratic trend model provided a more general incremental tendency projection, while the ARIMA model took into account for the details. For instance, the actual data show reduction in passenger volumes in 2003 and 2009. It can be seen that the ARIMA model followed the exact trend including the decrease situation but with a delay.

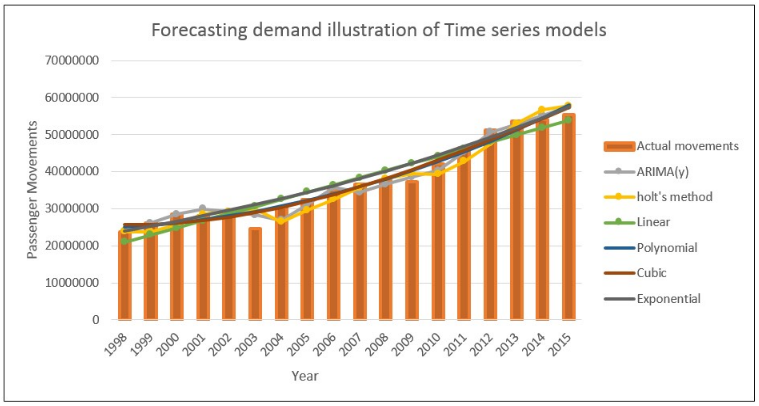

Fig. 5. Forecasting results from 2016 to 2033

Forecasting

Since the ARIMA (3, 2, 2) model and the quadratic trend model outperform other models, 18 more years passenger demands are forecasted using these two models. Figure 5 demonstrates that from 2018 onwards, the quadratic trend model reveals a faster growth rate.

The phenomenon occurs may be because the quadratic trend model implies an increasing growth rate of air passenger demand. On the other hand, the ARIMA model took into consideration that there were decreasing cases in the period of 2004 and so on. Hence, there is possibility during next 18 years the air passenger movements may decrease due to external causes.

It is noticeable that air passenger movements are expected to be doubled by using the ARIMA model and tripled by using the quadratic trend model. Under such circumstances, there is necessity to plan for expansion of infrastructure and operational procedures properly and promptly to keep service standard. Moreover, accurate forecasting information can lead to appropriate timing for facility construction with minimum effect on service. In addition, the long-term forecasting also provides information for aircraft ordering and design in consideration of bigger aircraft to carry more passengers [20]. Simultaneously, the accurate forecast can contribute to identifying capacity constraints in advance [21]. Decisions of network planning and management, fleet assignment, man power planning, flight scheduling, and revenue managements and so on are also affected by forecasting information [22].

CONCLUSION AND RECOMMENDATIONS

This study examined the ability of several time-series models for predicting the yearly number of air passenger movements in Singapore. The results in this paper show that the ARIMA model was found to be the best in time-series models. The quadratic trend model was also found to be highly accurate for the passenger movements forecast. The next 18 years' (2016-2033) passenger volume was forecasted by using the quadratic trend model and the ARIMA (3, 2, 2) model.

It is expected in the future seasonality or cyclic can also be taken into account. Holt-Winters exponential smoothing and the Seasonal Autoregressive-Integrated-Moving Average Model (SARIMA) shall be employed for prediction.

Declaration of Conflicting Interests

There are no financial/no-financial conflicts of interest in the current study.

Acknowledgement

This research was sponsored by the ATMRI of NTU and CAAS via ATMRI Project No. 2014-D2-ZHONG for Regional Airspace Capacity Enhancement - ASEAN Pilot.

REFERENCES

[1] Skytrax. (2016). *Singapore Changi airport named as the world's best airport* [Online]. Available: https://goo.gl/ocH99x

[2] E. Onder and S. Kuzu, "Forecasting air traffic volumes using smoothing techniques," *Journal of Aeronautics and Space Technologies,* vol. 7, no. 1, pp. 65-85, 2014.

[3] V. Profillidis and G. Botzoris, "Air passenger transport and economic activity," *Journal of Air Transport Management,* 49, 23-27, 2015.

[4] J. D. Bermudez, J. V. Segura and E. Vercher, "Holt-Winters forecasting: An alternative formulation applied to UK air passenger data," *Journal of Applied Statistics,* vol. 34, no. 9, pp. 1075-1090, 2007.

[5] V. A. Profillidis, "Econometric and fuzzy models for the forecast of demand in the airport of Rhodes," *Journal of Air Trans-*

port Management, vol. 6, no. 2, pp. 95-100, 2000.
[6] R. A. Scarpel, "Forecasting air passengers at Sao Paulo international airport using a mixture of local experts model," *Journal of Air Transport Management,* vol. 26, pp. 35-39, 2013.
[7] E. Ma, Y. Liu, J. Li and S. Chen, "Anticipating Chinese tourists arrivals in Australia: A time series analysis," *Tourism Management Perspectives,* vol. 17, pp. 50-58, 2016.
[8] W. H. K. Tsui, H. O. Balli, A. Gilbey and H. Gow, "Forecasting of Hong Kong airport's passenger throughput," *Tourism Management,* vol. 42, pp. 62-76, 2014.
[9] H. Grubb and A. Mason, "Long lead-time forecasting of UK air passengers by Holt-winters methods with damped trend," *International Journal of Forecasting,* vol. 17, no. 1, pp. 71-82, 2001.
[10] A. Andreoni and M. N. Postorino, "A multivariate ARIMA model to forecast air transport demand," in *European Transport Conference, Association for European Transport and Contributors,* Strasbourg, France, Sep. 18, 2006, 1-14.
[11] Wikipedia. (2016). *Singapore Changi airport* [Online]. Available: https://goo.gl/TyUDLZ
[12] R. J. Hyndman and G. Athanasopoulos, *Forecasting: Principles and Practice.* OTexts, 2014.
[13] C. Chatfield, *Time-Series Forecasting.* Boca Raton, FL: CRC Press, 2000.
[14] R. R. Yager and N. Alajlan, "A note on mean absolute deviation," *Information Sciences,* vol. 279, pp. 632-641, 2014.
[15] M. Torabi and J. N. K. Rao, "Estimation of mean squared error of model-based estimators of small area means under a nested error linear regression model," *Journal of Multivariate Analysis,* vol. 117, pp. 76-87, 2013.
[16] C. D. Lewis, *Industrial and Business Forecasting Methods: A Practical Guide to Exponential Smoothing and Curve Fitting.* London, UK: Butterworth, 1982.
[17] J. Fox. (2002).*Nonlinear regression and nonlinear least squares* [Online]. Available: https://goo.gl/HSLaLj.
[18] K. Taneja, S. Ahmad, K. Ahmad and S. D. Attri, "Time series analysis of aerosol optical depth over New Delhi using Box-Jenkins ARIMA modeling approach," *Atmospheric Pollution Research,* vol. 7, no. 4, pp. 585-596, 2016.
[19] W. W. S. Wei, *Time Series Analysis.* Redwood City, CA: Addison-Wesley Publication, 1994.
[20] V. A. Profillidis, "An ex-post assessment of a passenger demand forecast of an airport," *Journal of Air Transport Management,* vol. 25, pp. 47-49, 2012.
[21] K. Kolker, P. Bieblich and K. Lutjens, "From passenger growth to aircraft movements," *Journal of Air Transport Management,* vol. 56, pp. 99-106, 2016.
[22] R. B. C. Benitez, R. B. C. Paredes, G. Lodewijks and J. L. Nabais, "Damp trend grey model forecasting method for airline industry," *Expert Systems with Applications,* vol. 40, no. 12, pp. 4915-4921, 2013.

— This article does not have any appendix. —