

Forecasting airport passenger traffic: the case of Hong Kong International Airport

Wai Hong Kan TSUI^{1*}, Hatice Ozer BALLI² and Hamish GOW²

¹School of Aviation, Massey University, New Zealand

²School of Economics, Massey University, New Zealand

Abstract: *Hong Kong International Airport is one of the main gateways to Mainland China and the major aviation hub in Asia. An accurate airport traffic demand forecast allows for short and long-term planning and decision making regarding airport facilities and flight networks. This paper employs the Box-Jenkins Autoregressive Integrated Moving Average (ARIMA) methodology to build and estimate the univariate seasonal ARIMA model and the ARIMX model with explanatory variables for forecasting airport passenger traffic for Hong Kong, and projecting its future growth trend from 2011 to 2015. Both fitted models are found to have the lower Mean Absolute Percentage Error (MAPE) figures, and then the models are used to obtain ex-post forecasts with accurate forecasting results. More importantly, both ARIMA models predict a growth in future airport passenger traffic at Hong Kong.*

Introduction

After the 1997 changeover in Hong Kong from Britain sovereignty to Mainland China control, and 14 years after the opening of Hong Kong International Airport (HKIA), airport passenger traffic has been in steady growth generally. However, several authors (Hobson & Ko, 1994; Seabrooke, Hui, La & Wong, 2003; Zhang, 2003) have predicted a decline in air passenger and cargo throughput for Hong Kong International Airport and have raised concerns regarding its dominant role as an international hub and a gateway to Mainland China. Furthermore, Zhang, Hui, Leung, Cheung and Hui (2004, p.95) stated that “neither the [Hong Kong’s] gateway role nor the hub role should be taken for granted, and it will be risky to think that the hub role may be maintained forever and ... high growth rates will persist for a long time”.

An accurate and reliable airport passenger demand forecasting is an integral component for short-term and long-term planning and decision making regarding airport infrastructure development and flight networks.

However, Hong Kong’s future airport passenger demand is somewhat less well researched, although tourist arrival forecasting has been recently studied (Cho, 2003), and thereby indicates a gap for this paper to fill. In addition, to my best knowledge, this paper is the first empirical study on airport passenger demand for HKIA using time-series methods. The objective of this paper is thus to build time-series forecasting models to accurately predict airport passenger traffic at Hong Kong, and making a projection regarding its growth in future air passenger traffic from 2011 to 2015.

* Correspondence author: Mr Kan TSUI, School of Aviation, Massey University, Turitea Campus, Private Bag 11222, Palmerston North 4442, New Zealand. <http://aviation.massey.ac.nz/>. Email Wai.Tsui.1@uni.masey.ac.nz

Methods

Sample

Discreet months between January 2001 and February 2011 comprise the ‘subjects’ of the research. This amounts to 122 observations (equal to 12 months times 10 years, plus two additional months). The forecasting models are estimated and built using the data from January 2001 to November 2010, and the remaining data from December 2010 to February 2011 are retained to assess the ex-post forecasting performance of the models.

The variables of interest for our research are several, such as “monthly airport passenger traffic data for HKIA” (obtained from the Hong Kong Airport Authority), “GDP per capita”, “monthly visitors by air”, “connecting traffic”, etc. Of these variables, the “monthly airport passenger traffic data for HKIA” (obtained from the Hong Kong Airport Authority) will act as our main dependent variable, whose future demand we want to project into the future. The remaining variables, collected from the Census and Statistics Department of Hong Kong, the International Monetary Fund and the Hong Kong Tourism Board, will act as our main explanatory variables.

Statistical analyses

This paper employs the Box-Jenkins univariate seasonal ARIMA model and ARIMAX model for predicting future trends in airport passenger traffic for HKIA from March 2011 to December 2015 (Uddin, McCullough & Crawford, 1985; Cheung, 1991); Kawad & Prevedouros, 1995; Prevedouros, 1997; Chen & Chen, 2003; Andreoni & Postorino, 2006; Jia, Sun, Wan & Ma, 2007; Payne & Taylor, 2007; . Lee, 2009; Abdelghany & Guzhva, 2010; Samagaio & Wolters, 2010). The Box-Jenkins ARIMA methodology (1976) can be described as follows:

Theoretically, Box-Jenkins ARIMA models are built empirically from the observed time series relying on three underlying process components: Autoregressive (AR), Integrated (I) and Moving Average (MA) (Box & Jenkins, 1976; Pankratz, 1983; Box, Jenkins & Reinsel, 2008; Gujarati & Porter, 2009). In practice, the Box-Jenkins process uses a p th-order autoregressive process (p), a q th-order moving average process (q), and a level of differencing (d) to build the most appropriate fitted model for forecasting time series.

The autoregressive process implies that an observed event at a specific time period depends on its value in the previous time period and error term. The moving average process means that an observed event at a specific time is dependent on the previous values of its error term. The integrated process indicates the level of differencing required to make time series stationary. Most importantly, the ARIMA models can only be used when the time series is stationary.

The general Box-Jenkins ARIMA model is considered as the non-seasonal ARIMA (p,d,q) model. However, the seasonality is very common with air passenger traffic or tourist arrival time series. Therefore the seasonal ARIMA (P,D,Q)_s model is developed taking into account seasonal patterns in the time series, such as quarterly, semi-annual or annually. Combining the non-seasonal ARIMA (p,d,q) model and the seasonal ARIMA (P,D,Q)_s model, and then the seasonal ARIMA can be expressed as follows: SARIMA (p,d,q) \times (P,D,Q)_s⁴

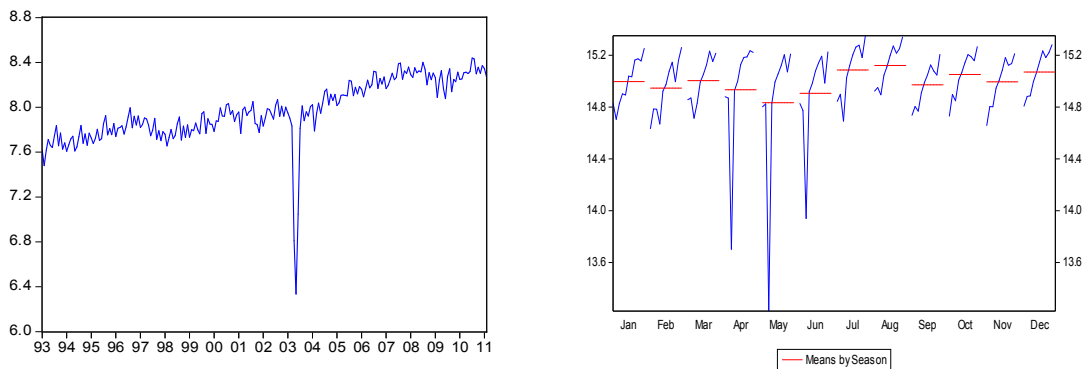
The Box-Jenkins ARIMA model thus includes four specification techniques: identification, estimation, diagnostic checking, and forecasting (Box & Jenkins, 1976).

⁴ p and P denotes non-seasonal and seasonal autoregressive process; q and Q denotes non-seasonal and seasonal moving average process; d and D denotes level of differencing required to make time series being stationary for non-seasonal and seasonal process.

Results

Figure 1 displays the monthly log-airport passenger traffic passing through Hong Kong between January 2001 and February 2011.⁵ Based on graphical analysis, the time series exhibits an upward trend along with the possibility of seasonal patterns. More importantly, the SARS outbreak caused declines in airport passenger traffic between late 2002 and mid-2003. In the seasonality plot, we can observe a peak every July and August, indicating high travel periods during the summer holidays, followed by a drop in September. Another peak occurs every December and March.

Figure 1: Log-airport passenger traffic (left-side) and the seasonality plot (right-side)



Before estimating log-airport passenger traffic time series, we checked whether the time series data was stationary. The Augmented Dickey-Fuller test was significant ($t = -4.376$, $p < .01$), indicative of the time series being stationary. Thus, further differencing process were not required.

Moreover, the correlograms of Autocorrelation (ACF) and Partial Autocorrelation (PACF) indicate that the time series have some seasonal, autoregressive and moving average processes. This suggests the use of the SARIMA model. After extensive trial-and-error, using the lowest test results of Akaike Information Criterion (1974) and Schwarz Information Criteria (1978), we decided that the best-fit model with best forecasting performance for predicting airport passenger traffic for Hong Kong was the univariate seasonal ARIMA $(1,0,1) \times (1,0,1)_{12}$ model.

Furthermore, we also found that the residuals fell within the 95% of confidence level in the residual correlograms of ACF and PACF,⁶ and also there are “white noise” process.⁷ In addition, the Ljung-Box Q -statistics (1978) suggests one fails to reject the null hypothesis of no autocorrelation. After performing the Ordinary Least Squares (OLS) estimation procedure, Table 1 shows all of the AR and MA terms are statistically significant at above the 99% confidence level, as well as the fitted univariate seasonal

⁵ The variance appears significantly change much over time in the original time series of monthly airport passenger traffic, and then the logarithm transformation is chosen stabilise the variance.

⁶ For the sake of brevity, the correlograms of residuals of univariate seasonal ARIMA model are not reported in this paper. The correlograms would be available upon request.

⁷ “White noise” process of residuals means that they are independent and identically distributed with zero mean and variance σ^2 ($\epsilon_i \sim \text{iid } N(0, \sigma^2)$).

ARIMA model has the overall predictable power with an adjusted- R^2 of 0.850 and it is quite accurate with a small MAPE figure (0.633%) (Lewis, 1982).

Table 1: Estimated results of the univariate seasonal ARIMA model

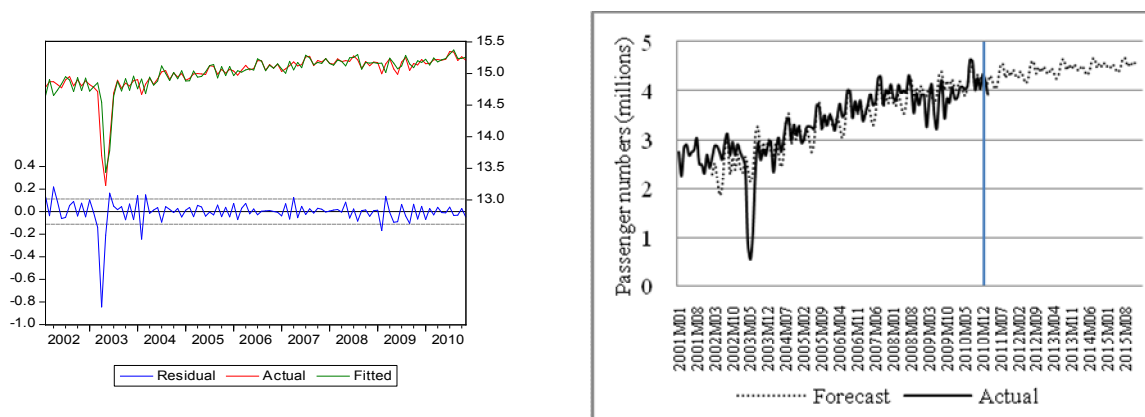
| Dependent variable = log-airport passenger numbers | | |
|--|-------------|--------------|
| Explanatory variables | Coefficient | t-statistics |
| Constant | 15.693*** | 17.299 |
| Trend | -0.001 | -0.287 |
| AR(1) | 0.574*** | 7.090 |
| SAR(12) | 0.858*** | 9.366 |
| MA(1) | 0.650*** | 5.382 |
| SMA(12) | -0.944*** | -34.152 |
| R^2 | 0.858 | |
| Adjusted- R^2 | 0.850 | |
| MAPE (%) | 0.633 | |

Remarks:

- 1) Method: Ordinary Least Squares (OLS)
- 2) Sample (adjusted): 2002M02–2010M11
- 3) Included observations: 106 after adjustments
- 4) *** indicates the variable is significant in either at above the 99% confidence level (two-tailed)

In addition, Figure 2 shows the comparison of residuals (the difference between the forecasted airport passenger traffic with actual values). Two main conclusions are derived from the residual patterns. First of all, the fitted values are sufficiently close to the actual values; particularly the graph captures the adverse impact of SARS outbreak upon airport passenger traffic at Hong Kong. Second, the fitted univariate seasonal ARIMA model has the ability to preserve both short and long-term memory and can be used for long-term forecast of passenger traffic for HKIA. In addition, HKIA's future passenger traffic demand is projected to grow towards 2015.

Figure 2: Actual, fitted and residuals for the univariate seasonal ARIMA model (left-side) & Monthly passenger traffic projection (right-hand)



After selecting the univariate seasonal ARIMA model, this paper continues to apply the ARIMAX model (i.e., the multivariate ARIMA model). In the essence of ARIMAX

modelling, the forecasting of future airport passenger traffic demand for HKIA can further be explained by the changes in originating or local traffic, visitors by air, connecting traffic (transfer and transit passenger traffic) via the airport,⁸ along with other interventions or shocks (i.e., the SARS outbreak, Cross-Strait direct flight agreement between China and Taiwan, fuel price, Individual Travel Scheme). In order to estimate the ARIMX model, future forecasted values relate to the explanatory variables and the likely impacts of interventions (shocks) need to be estimated. This paper resorts to a combination of available forecasts/estimates from the external sources and the ARIMA application. For the originating traffic, the forecasts of Hong Kong's GDP per capita obtained from the International Monetary Fund (IMF) are used as the proxy due to the direct correlation between airport demand and GDP.⁹ Unfortunately, future forecasts for visitors by air and connecting traffic are more difficult to obtain due to the lack of published data, and thus their future forecasted values are predicted by employing the Box-Jenkins ARIMA methodology. After extensive trial-and-error specification, the best-fit models for log-visitors by air and log-connecting traffic are the SARIMA (4,0,2)×(1,0,0)₁₂ model and the SARIMA (2,0,2)×(1,0,0)₁₂ model, respectively. Both fitted SARIMA models have the MAPE figures of 1.091% and 0.841%.

Moreover, the likely effects of interventions (shocks) are captured by employing dummy variables in accordance with their permanent or temporary effects.¹⁰ For example, fuel price is assumed to maintain at the level of more than US\$80 per barrel between March 2011 and December 2015.

$$Fuel_t = \begin{cases} 1, & \text{if } t \geq \text{US\$80 per barrel (at and after the intervention)} \\ 0, & \text{otherwise} \end{cases}$$

As described above, the ARIMAX modelling is first to incorporate the selected univariate SARIMA (1,0,1)×(1,0,1)₁₂ model, and then we employed the General-to-Specific (GS) approach to determine the lags of log-GDP per capita, log-visitors by air and log-connecting traffic by eliminating variables which are statistically insignificant, as well as inserting identified interventions (shocks) during the regression analysis (Henry, 1995; Song, Wong & Chon, 2003; Balli & Elsamadisy, 2010). Table 2 shows the parameters of regression coefficients and the ARIMAX model has a good fit with an adjusted- R^2 of 0.993 and the lower MAPE figure of 0.127%.

Given no further improvement can be achieved for the ARIMAX model, and also the ACF and PACF correlograms and the Ljung-Box Q -statistics (1978) confirm the “white noise” characteristics of residuals. From Figure 3, we found that the fitted airport passenger traffic for HKIA are more closely resemble the actual values shown, suggesting the ARIMAX model has more predictive power with additional explanatory

⁸ Hong Kong International Airport experiences more transfer or transit passenger traffic, since it plays a role of the international gateway airport to the Pearl River Delta region and Mainland China, and also the international transportation hub to Asia and Southeast Asia (O’Conner 1995; Oum & Yu, 2000; Zhang *et al.*, 2004; Mason, 2007).

⁹ For checking purpose, IMF and the Census and Statistics Department of Hong Kong offered the same figures of Hong Kong’s GDP per capita between 2001 and 2010.

¹⁰ For the sake of brevity, the future forecasts for log-visitors by air and log-connecting traffic, the assumptions of future impacts of intervention (shocks) are not reported. These figures would be available upon request.

variables than the univariate seasonal ARIMA model. Similarly, the ARIMX model also predicts a steady growth in future airport passenger traffic at Hong Kong for the period from 2011 to 2015.

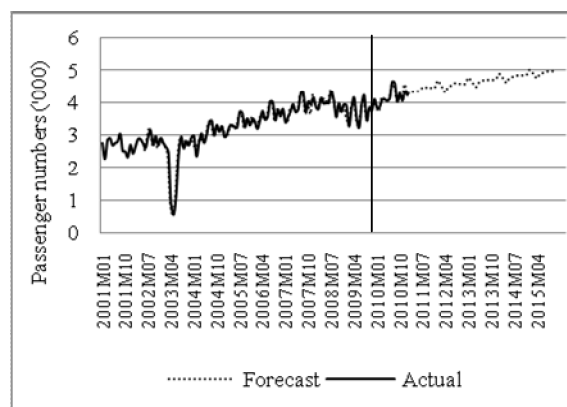
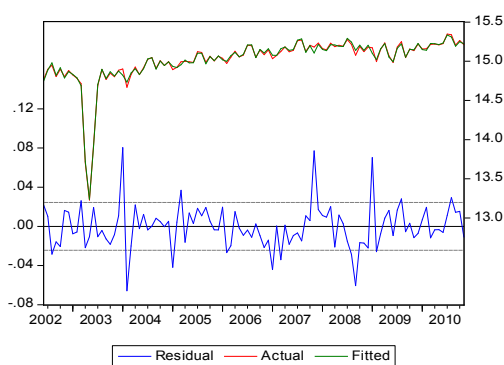
Table 2: Estimated results of the ARIMAX model

| Dependent variable = log-airport passenger traffic | | |
|--|-------------|--------------|
| Explanatory variable | Coefficient | t-statistics |
| Constant | 6.095*** | 7.995 |
| Trend | 0.002*** | 3.312 |
| Log-GDP per capita | 0.350* | 1.891 |
| Log-GDP per capita(-1) | -0.198 | -0.628 |
| Log-GDP per capita(-2) | -0.326* | -1.875 |
| Log-GDP per capita(-3) | -0.371*** | -2.986 |
| Log-GDP per capita(-4) | 0.404*** | 3.602 |
| Log-Connecting traffic | 0.237*** | 5.416 |
| Log-visitors by air | 0.489*** | 13.998 |
| SARS | -0.027** | -2.085 |
| Cross Strait | -0.028*** | -3.674 |
| Fuel price | 0.050*** | 5.292 |
| IVS | -0.013 | -0.876 |
| AR(1) | 0.482*** | 3.062 |
| SAR(12) | 0.789*** | 13.037 |
| MA(1) | -0.997*** | -35.981 |
| SMA(12) | -0.931*** | -41.886 |
| R ² | 0.994 | |
| Adjusted-R ² | 0.993 | |
| MAPE(%) | 0.127 | |

Remarks:

1. Method: Ordinary Least Squares (OLS)
2. Sample (adjusted): 2002M06–2010M11
3. * and ** and *** indicates the variable is significant in either at above the 90%, 95% and 99% confidence level (two-tailed)
4. The lags of each explanatory variable is decided when it become statistically significant using *t*-statistics
5. AR and MA terms included are to capture autoregressive and moving average relationships in the time series

Figure 3: Actual, fitted and residuals for the ARIMAX model (left-side) & Monthly passenger traffic projection (right-hand)



Evaluation of forecasts

It is important to check the forecast accuracy of both fitted ARIMA models by evaluating ex-post forecasts using the period of December 2010 and February 2011. After transforming log-airport passenger traffic to absolute values, Table 3 shows the forecasting performance of both fitted models by comparing their actual airport passenger traffic and forecasted values.¹¹ We found that the forecasted errors of univariate seasonal ARIMA model are smaller compared to those of ARIMX model. Another finding is that, for both models, the forecasted errors are very accurate within two months forecast horizon, and the forecasted errors increase remarkably when the forecast horizon gets to three month. This indicates that both models may experience larger forecasted errors for long-term forecasts.

Table 3: Forecast performance between ARIMA models

| Periods | Univariate seasonal ARIMA model | | | ARIMAX model | |
|---------|------------------------------------|-----------|--------------------------|--------------|--------------------------|
| | Actual | Forecast | Forecasted errors (%) | Forecast | Forecasted errors (%) |
| 2010M12 | 4,328,138 | 4,353,869 | -0.59 | 4,562,261 | -5.41 |
| 2011M01 | 4,206,217 | 4,220,524 | -0.34 | 4,243,759 | -0.89 |
| 2011M02 | 3,909,319 | 4,162,938 | -6.49 | 4,317,438 | -10.44 |

Conclusions

In summary, the best-fit Box-Jenkins univariate seasonal ARIMA model and the ARIMAX model are built for predicting airport passenger traffic for Hong Kong. The ex-post forecasts indicate that both fitted models provide accurate and reliable forecast results ranging from -0.59 to -10.44 percent. For the long-term forecasts, both models may suffer from larger forecasted errors. However, Hong Kong airport's future passenger traffic is projected to maintain a growth trend for the period of March 2011 to December 2015, and more importantly, this projection highlights the challenges for policy makers, airport authority and airline management to meet the increasing demand of airport passenger traffic at Hong Kong.

References

- ABDELGHANY A & VS GUZHVA (2010).** *A time series modelling approach for airport short-term demand forecasting.* Airport Management, 5(1), 72-87.
- AKAIKE H. (1974).** *A new look at the statistical model identification.* IEEE Transaction on Automatic Control, 19, 716-723.
- ANDREONI A & MN POSTORINO (2006).** *A multivariate ARIMA model to forecast air transport demand.* Proceedings of Association of European Transport 2006, France, September 18-20, 2006.
- BALLI F & EM ELSAMADISY (2010).** *Modeling the currency in circulation for the state of Qatar.* Munich Personal RePEc Archive (MPRA). Retrieved from <http://mpr.a.ub.uni-muenchen.de/20159/>

¹¹ See Wooldridge (2009), the procedures for transforming forecasted log-airport passenger traffic to forecasted absolute values should take into account of the residuals presented in the time series.

- BOX G E P & GM JENKINS (1976).** *Time series analysis: Forecasting and control.* San Francisco: Holden Bay.
- BOX G E P & GM JENKINS & GC REINSEL (2008).** *Time series analysis: Forecasting and control (4th eds.,).* New Jersey: John Wiley and Sons, Inc.
- CHEUNG CP (1991).** *Multivariate time series analysis on airport transportation.* Unpublished master's thesis, The University of Hong Kong.
- CHO V (2003).** *A comparison of three different approaches to tourist arrival forecasting.* *Tourism Management*, 24, 323-330.
- DICKEY DA & FULLER WA (1976).** *Distribution of the estimates for autoregressive time series with a unit root.* *Journal of the American Statistical Association*, 74, 427-431.
- GUJARATI DN & PORTER DC (2009).** *Basic Econometric (5th eds.,).* New York: McGraw-Hill.
- HENRY DF (1995).** *Dynamic econometrics: an advanced text in econometrics.* Oxford: Oxford University Press.
- HOBSON J S P & G KO (1994).** *Tourism and Politics: The Implications of the Change in Sovereignty on the Future Development of Hong Kong's Tourism Industry.* *Journal of Travel Research*, 32(2), 1-8.
- JIA CL, SUN Y, WANG L B & MA Y L (2007).** *Fitting analysis of the airport passenger throughput based on ARIMA model.* Proceedings of the 6th International Conference on Management, Wahan: August 3-5, 2007.
- KAWAD S & PD PREVEDOUROS (1995).** *Forecasting air travel arrivals: Model development and application at the Honolulu International Airport.* Transportation Research Board, 1506, 18-25.
- LEE W T (2009).** *The study of air traffic forecasting model – the Southern International Airport.* Unpublished master's thesis (in Chinese): National University of Kaohsiung.
- LEWIS CD (1982).** *International and business forecasting methods.* London: Butterworths.
- LIM C & MCALEER M (2002).** *Time series forecasts of international travel demand for Australia.* *Tourism Management*, 23, 389-396.
- LIM C, MCALEER M, & MIN J C H (2009).** *ARMAX modelling of international tourism demand.* *Mathematics and Computers in Simulation*, 79, 2879-2888.
- JUNG GM & G E P BOX (1978).** *On a measure of lack of fit in time series models.* *Biometrika*, 65(2), 297-303.
- MASON K (2007).** *Is the "Gateway" Concept Useful or Relevant to the Passenger Aviation Market?* United Kingdom: Cranfield University Press.
- O'CONNOR K (1995).** *Airport development in Southeast Asia.* *Journal of Transport Geography*, 3(4), 269-279.
- OUM TH & YU C (2000).** *Shaping air transport in Asia Pacific.* Aldershot: Ashgate.
- PANKRATZ A (1983).** *Forecasting with univariate Box-Jenkins Models: Concepts and cases.* New York: John Wiley & Sons, Inc.
- PAYNE JE & JP TAYLOR (2007).** *Modelling and forecasting airport passengers: a case study for an introductory forecasting course.* *International Journal of Information and Operations Management Education*, 2(2), 167-182.
- PREVEDOUROS PD (1997).** *Origin-Specific visitor demand forecasting at Honolulu International Airport.* Transport Research Board, 1600, 18-27.
- SAMAGAIO A & M WOLTERS (2010).** *Comparative analysis of government forecasts for the Lisbon airport.* *Journal of Air Transport Management*, 16, 213-217.

SCHWARZ GE (1978). *Estimating the dimension of a model.* Annals of Statistics, 6, 461-464.

SEABROOKE W, HUI E C M, LAM W H K & WONG G K C (2003). *Forecasting cargo growth and regional role regional role of the port of Hong Kong.* Cities, 20(1), 51-64.

SONG H, WONG K K F & CHON K K S (2003). *Modelling and forecasting the demand for Hong Kong tourism.* Hospitality Management, 22, 435-451.

WILLIAMS BM (2007). *Multivariate vehicular traffic flow prediction: Evaluation of ARIMAX modelling.* Transportation Research Board, 1776, 194-200.

UDDIN W, BF MCCULLOUGH & MM CRAWFORD (1985). *Methodology for forecasting air travel and airport expansion needs.* Transport Research Board, 1025, 7-14.

WOOLDRIDGE JM (2009). *Introductory Econometrics: A modern approach (4th eds.).* US: South-Western Cengage Learning.

ZHANG A (2003). *Analysis of an international air-cargo hub: the case of Hong Kong.* Journal of Air Transport Management, 9, 123-138.

ZHANG A, HUI G W L, LEUNG L C, CHEUNG W & HUI V Y (2004). *Air Cargo in Mainland China and Hong Kong.* England: Ashgate.