Time Series Modelling of Tourist Arrivals to Malaysia

Mahendran Shitan

Department of Mathematics,
Faculty of Science,
and
Applied & Computational Statistics Laboratory,
Institute for Mathematical Research,
Universiti Putra Malaysia

Abstract

Tourism comprises of the activities of persons traveling to and staying in places outside their usual environment for not more than one consecutive year. Tourism is one of the major foreign exchange earners in Malaysia. Therefore, forecasting tourist arrivals becomes important because it would enable tourist related industries like airlines, hotels, food and catering services, etc., to plan and prepare their activities in an optimal way.

Previous studies have suggested various time series models for modelling monthly tourist arrivals to Malaysia. Since then, new observations have become available and it is important to update these models.

Therefore, in this paper we update and compare the performance of three time series models for modelling tourist arrivals to Malaysia. One of them is within the class of ARMA models and the other two are in the class of ARFIMA models.

Keywords: Time Series, Forecasting, ARMA, ARFIMA, Tourist Arrivals

1 Introduction

According to the definition by the World Tourism Organization (WTO)/United Nations recommendations on Tourism Statistics, tourism comprises the activities of persons travelling to and staying in places outside their usual environment for not more than one consecutive year for leisure, business and other purposes. In the annual report of WTO in Asia and the Pacific, 2000, the statistics indicated that the tourism industry has grown at a very accelerated pace over the latter half of the past century. The projections for continued growth make it an important sector of the world economy which is growing faster than the world trade.

Over the last decade, the Asia Pacific region (Amrik Singh, 1997) which includes North-East Asia, South-East Asia and Oceania, has been the fastest growing tourism regions in the world. Tourism receipts has increased steadily from RM 4500 million in 1990 to RM11251 million in 1996 (Annual Statistical Tourism Report, 1998). Tourism is said to be the world's largest industry in the World Travel and Tourism Report (WTTC) as the industry is not only able to contribute to the employment sector in creating job opportunities but its contribution is also significant in generating wealth of a country. Therefore, tourism research in various fields like promoting, marketing, forecasting and planning is indeed necessary and important. With the ability in forecasting tourist arrivals, it helps the Government and other tourism related organizations in their future planning.

In a previous research (Chuah, 2001) on Forecasting Tourist Arrivals to Malaysia, various forecasting techniques like the Integrated Autoregressive Moving Average (ARIMA) forecasting method, the naïve method, simple exponential smoothing method, Holt's linear method and Holts-Winter's trend and seasonal method were considered. It was found that the Moving Average 10 or MA(10) model was the best forecasting model based on the smallest root mean square error (RMSE). In a subsequent research (Mahendran Shitan and Pauline Mah Jin Wee, 2003a) a comparison was made to investigate if the ARAR and the long memory (ARFIMA) models perform better than the previously proposed MA (10) model. In another study (Mahendran Shitan and Pauline Mah Jin Wee, 2003b) comparison of ARFIMA model and Regression model with ARMA errors were considered and the ARFIMA(7,0.42,10) model emerged better. Sudden intervening events could also immediately or eventually affect tourist arrivals and in view of this Intervened Time Series Models have been considered (see Mahendran Shitan and Tan Yiing Fei, 2004).

However, since 2004 fresh observations have become available and it is important to update previously proposed models. With that in mind, we undertake the time series modelling of tourist arrivals to Malaysia and in this paper we propose and compare the performance of three time series models. The time series models considered are within the class of ARMA and ARFIMA models.

In Section 2, the methodology of this research is discussed followed by the results in Section 3. Finally, the conclusions are contained in Section 4.

2 Methodology

In section 2.1 the data set is discussed and in section 2.2 the time series models used in this study are briefly explained.

2.1 The Data Set

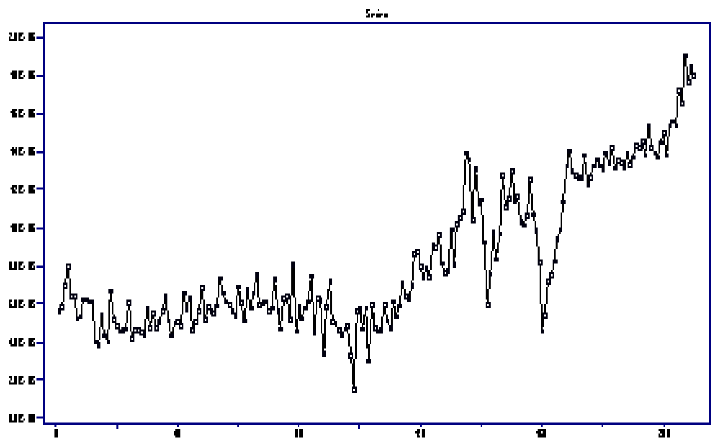

Figure 1: The time series plot of the monthly tourist arrivals data from January 1990 to June 2007.

The statistics of tourist arrivals to Malaysia is collected by the Immigration Department of Malaysia, Kuala Lumpur through the embarkation and disembarkation cards at all major exit points of Malaysia. The monthly tourist arrivals data from January 1990 to December 2000 that consists of

132 observations was obtained from the previous research project (Chuah, 2001) while the remaining 78 observations from January 2001 to June 2007 was accessed from the website of the Malaysia Tourism Promotion Board.

The time series plot from January 1990 to June 2007 consisting of a total of 210 monthly observations is shown in Figure 1. From the plot, it is clear that tourist arrivals has generally increased over the years and obviously it is not a stationary time series. There also appears to have some sort of seasonal pattern in it. There are also some unexpected dips and some events may have contributed to a drop in tourist arrivals at these points of time.

2.2 ARIMA and ARFIMA Models

The models considered in this study are the ARIMA and ARFIMA time series models. Therefore, we provide here a brief description of these two models. For a complete description, please refer to Brockwell and Davis (2002).

A zero mean stationary ARMA(p, q) process is defined as a sequence of random variables $\{X_t\}$ given by,

$$X_t - \phi_1 X_{t-1} - \ldots - \phi_p X_{t-p} = Z_t + \theta_1 Z_{t-1} + \ldots + \theta_q Z_{t-q},$$

where $\{Z_t\}$ is a sequence of uncorrelated random variables with zero mean and constant variance, denoted as $\{Z_t\} \sim WN\,(0, \sigma^2)$. We define $\{X_t\}$ to be an ARMA(p, q) with mean μ if $\{X_t - \mu\}$ is an ARMA(p, q) process.

If d is a non-negative integer, then $\{X_t\}$ is said to be an ARIMA(p, d, q) process if $(1-B)^d X_t$ is a ARMA(p, q) process, where B is the usual backward shift operator.

On the other hand if $0 < |d| < 0.5$, we obtain what is known as a long memory process or a Fractionally Integrated ARMA, [ARFIMA (p, d, q)] process. The ARFIMA processes satisfy the difference equation,

$$(1 - B)^d \phi\,(B)\,X_t = \theta\,(B)\,Z_t,$$

where $\{Z_t\} \sim WN\,(0, \sigma^2)$, $\phi\,(z) = 1 - \phi_1 z - ... - \phi_p z^p$ satisfying $\phi\,(z) \neq 0$ and $\theta\,(z) = 1 + \theta_1 z + ... + \theta_q z^q$, satisfying $\theta\,(z) \neq 0$ for all z such that $|z| \leq 1$. The ARFIMA process is a stationary process with much more slowly decreasing autocorrelation function (see Brockwell and Davis, 2002 for further details).

3 Results

In this section we present the results of our study.

We used the data from January 1990 to June 2007 (210 observations) for the modelling process and used the data from January 2005 to June 2007 (30 observations) to evaluate the forecasting performance of our fitted models.

Let $\{Y_t\}$ represent the monthly tourist arrivals to Malaysia. After an extensive search amongst many possible models, we narrowed down to the following models.

Model 1

$$\begin{aligned}(1-B)(1-B^{12})Y_t - 3001 &= Z_t - 0.1590Z_{t-1} - 0.06846Z_{t-2} + \\ & 0.004975Z_{t-3} - 0.07361Z_{t-4} + \\ & 0.05964Z_{t-5} + 0.2042Z_{t-6} - \\ & 0.1454Z_{t-7} + 0.01290Z_{t-8} - \\ & 0.03701Z_{t-9} - 0.008276Z_{t-10} + \\ & 0.06926Z_{t-11} - 0.7976Z_{t-12}, \qquad (1)\end{aligned}$$

where $\{Z_t\} \sim WN(0, 11347500000)$.

Notice that $(1-B)(1-B^{12})Y_t - 3001$ is a ARMA(0, 12) model.

Model 2

$$\begin{aligned}(1-B)^{-0.2771}((1-B^{12})Y_t - 61750) &= Z_t + 1.042Z_{t-1} + 1.052Z_{t-2} + \\ & 1.089Z_{t-3} + 0.9828Z_{t-4} + \\ & 0.9369Z_{t-5} + 0.8849Z_{t-6} + \\ & 0.7433Z_{t-7} + 0.6870Z_{t-8} + \\ & 0.6505Z_{t-9} + 0.6716Z_{t-10} + \\ & 0.7118Z_{t-11}, \qquad (2)\end{aligned}$$

where $\{Z_t\} \sim WN(0, 14187700000)$.

Observe that $(1-B^{12})Y_t - 61750$ is an ARFIMA(0, −0.2771, 11) model.

Model 3

$$\begin{aligned}(1-B)^{-0.2058}((1-B)(1-B^{12})Y_t-3001) \;=\; & Z_t-0.07336Z_{t-1}+\\ & 0.02973Z_{t-2}+0.05749Z_{t-3}-\\ & 0.07138Z_{t-4}+0.1092Z_{t-5}+\\ & 0.1618Z_{t-6}-0.04193Z_{t-7}-\\ & 0.009354Z_{t-8}+0.003057Z_{t-9}+\\ & 0.01690Z_{t-10}+0.02465Z_{t-11}-\\ & 0.5496Z_{t-12}, \qquad (3)\end{aligned}$$

where $\{Z_t\} \sim WN(0, 13877800000)$.

Notice here that $(1-B)(1-B^{12})Y_t - 3001$ is an ARFIMA(0, −0.2058, 12) model.

The process of model fitting was done using the computer software ITSM2000.

Next, in this study we forecasted the next 30 values (January 2005 to June 2007) using these three models and compared it with the actual observations. Plots of the predicted values and actual observed values for Models 1, 2 and 3 are shown in Figures 2, 3 and 4 respectively.

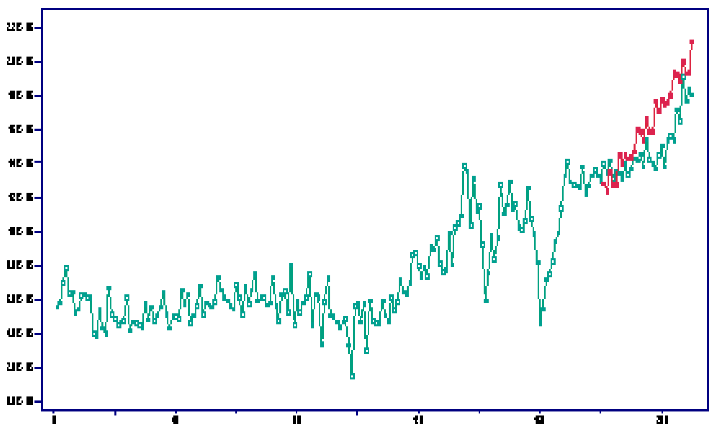

Figure 2: Plot of the predicted values and actual observed values using Model 1.

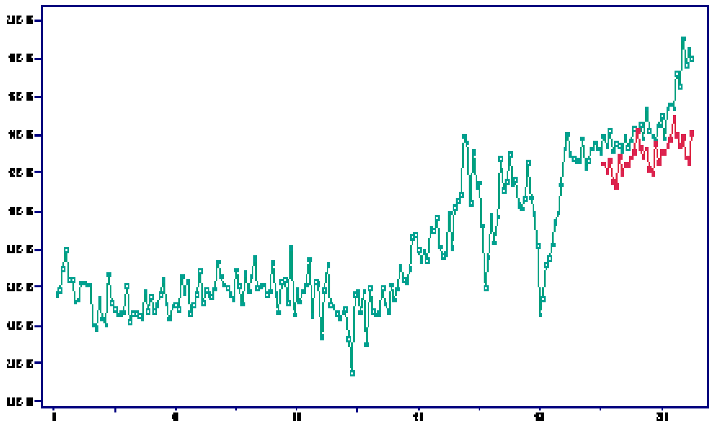

Figure 3: Plot of the predicted values and actual observed values using Model 2.

Clearly we can see that the forecasts produced by Model 3 is closer to the actual observed values when compared with Model 1 or 2. The forecast plot using Model 3 holds on tightly to the plot of the observed values. On the other hand, Model 1 somewhat over predicts while Model 2 under predicts.

Futher in order to evaluate the forecasting performance of these three models, the criteria chosen to measure the accuracy of the forecast in this

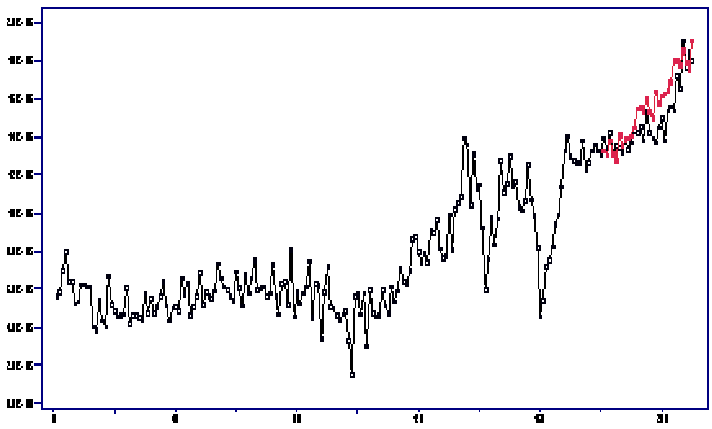

Figure 4: Plot of the predicted values and actual observed values using Model 3.

study are the mean absolute error (MAE), the root mean square error (RMSE) and the mean absolute percentage error (MAPE) which are given respectively by the following equations,

$$\text{MAE} = \frac{\sum_{i=1}^{n} |y_i - \hat{y}_i|}{n}, \quad \text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{n}}, \quad \text{MAPE} = \frac{\sum_{i=1}^{n} \left|\frac{y_i - \hat{y}_i}{y_i}\right|}{n} \times 100\%,$$

where y_i and $\hat{y}_i$ are the actual observed values and the predicted values respectively while n is the number of predicted values.

This observation is also evident from the computed values of MAE, RMSE and MAPE for models 1, 2 and 3 which are tabulated in Table 1.

Table 1: MAE, RMSE and MAPE values

	MAE	RMSE	MAPE
Model 1	168,305	196,690	11.28%
Model 2	194,256	239,326	12.39%
Model 3	94,018	115,116	6.38%

Clearly from Table 1 we notice that by all three criteria, Model 3 has the smaller values and as such a better forecasting performance. Hence, we would propose that Model 3 is appropriate for modelling tourist arrivals to Malaysia.

As such based on Model 3, the forecasted tourist arrivals from July 2007 to December 2008 together with the 95% forecast intervals are shown in Table 2.

A plot of these forecasted values is shown in Figure 5. Clearly from Table 2 and the plot, the monthly tourist arrivals are expected to increase in the forthcoming months and is expected to reach about 2.3 million towards the end of the year 2008. The 95% confidence interval indicates that monthly tourist arrivals could exceed 3 million towards the end of 2008 but more likely than not it would remain in the vicinity of about 2.3 million arrivals. The

Table 2: Forecast of Tourist Arrivals from July 2007 to December 2008 using Model 3

Month	**Forecasted Value**	**95% Confidence Interval**
July 2007	1829200	(1536500, 2121800)
Aug	1899600	(1538800, 2260300)
Sept	1842600	(1430100, 2255200)
Oct	1936500	(1476900, 2396100)
Nov	1968000	(1478400, 2457600)
Dec	1961700	(1434300, 2489100)
Jan 2008	2099100	(1522900, 2675300)
Feb	2030200	(1419200, 2641100)
Mar	2200500	(1559600, 2841500)
Apr	2096400	(1428100, 2764700)
May	2119800	(1425200, 2814500)
Jun	2121400	(1400800, 2842100)
July	2151500	(1363800, 2939200)
Aug	2226800	(1395600, 3058100)
Sept	2174400	(1302100, 3046800)
Oct	2272500	(1358800, 3186200)
Nov	2307900	(1363000, 3252800)
Dec	2305400	(1322700, 3288100)

lower confidence interval indicates that monthly tourist arrivals would not fall below the 1.3 million mark at the end of 2008.

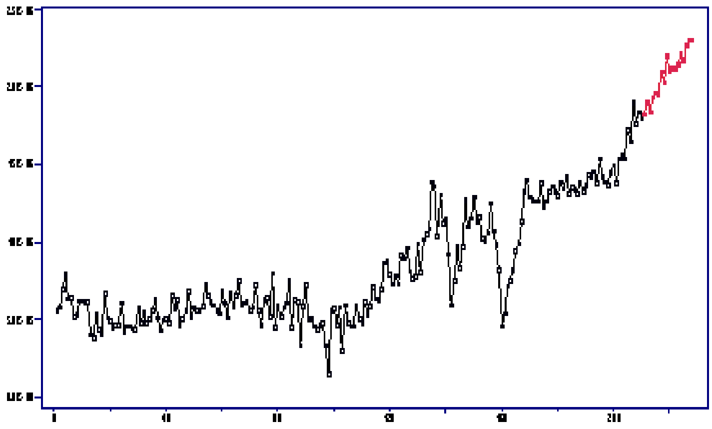

Figure 5: Forecast for July 2007 to December 2008 using Model 3.

4 Conclusion

The objective of this paper was to update existing time series models for forecasting tourist arrivals to Malaysia as fresh observations have become available.

We found that Model 3 may serve as a reasonable model since the MAE, MAPE and RMSE values had the smallest forecast errors when compared with the other two models. Based on Model 3, monthly tourist arrivals are expected to increase steadily and could possibly reach about 2.3 million at the end of 2008. The approximate 95% forecast interval for monthly tourist arrivals, when rounded is about 1.3 to 3.3 million, at the end of the year 2008.

This forecast is of course based on the assumption that events, like government policy, promotion campaigns, natural or man made events, etc., do not change drastically. For further research, some explanatory variables which may also influence tourist arrivals, like exchange rates, promotion campaigns, government policy changes could be included.

References

Amrik Singh (1997), Asia Pacific Tourism Industry: Current Trends And Future Outlook, *Asia Pacific Journal of Tourism Research.*

http://www.hotel-online.com/Trends/AsiaPacificJournal/

AsiaPacificTourismOutlook_1997.html. Accessed on 4 July 2002.

Brockwell, P.J. and Davis, R.A. (2002), *Introduction To Time Series And Forecasting*, 2nd Edition, Springer-Verlag, New York.

Chuah, S.C. (2001), *Forecasting Tourist Arrivals To Malaysia*, Masters Project, Universiti Putra Malaysia, Malaysia.

Mahendran Shitan and Pauline Mah Jin Wee (2003a), ARAR and Long Memory Modelling of Tourist Arrivals to Malaysia, *Proceedings of National Conference on Management Science and Operations Research,* Vol. 1, 258-271, 24-25 June, 2003.

Mahendran Shitan and Pauline Mah Jin Wee, (2003b), A comparison between ARFIMA Model and Regression Model with ARMA Errors for Tourist Arrivals in Malaysia, *Proceedings of National Seminar on Decision Science,* 109-114, 15-16 October, 2003, Langkawi, Kedah.

Mahendran Shitan and Tan Yiing Fei, (2004), A Comparative Study between ARFIMA Models and Intervened Time Series Model for Tourist Arrivals to Malaysia. *Paper presented at the Asia Pacific and Business Conference 2004*, 7-9 October, 2004, Penang

Malaysia Tourism Promotion Board (2001), Statistics: Malaysia Tourist Arrivals 2001. http://www.tourism.gov.my/statistic/2001.htm. Accessed on 23 May 2002.

Malaysia Tourism Promotion Board (2002), Statistics: Malaysia Tourist Arrivals, January to June 2002. http://www.tourism.gov.my/statistic/2002.htm. Accessed on 24 August 2002.

Malaysia Tourism Promotion Board (2004), Statistics: Malaysia Tourist Arrivals, July2002 to Feb 2004. http://www.tourism.gov.my/statistic/statiscs.asp. Accessed on 13 April 2004.

Malaysia Tourism Promotion Board (2007), Statistics: Malaysia Tourist Arrivals. http://www.tourism.gov.my/statistic/statiscs.asp. Accessed on 1 September 2007.

World Tourism Organization, Methodological Notes, Concepts and Definitions.

http://www.worldtourism.org/market_research/facts&figures/menu.htm. Accessed on 5 August 2002.

—— (1996-1998), *Annual Statistical Tourism Report*, Research Division, Malaysia Tourism Promotion Board.

—— (2002), Malaysia Tourism, *A bimonthly publication of Tourism Malaysia*, March-April, Tourism Malaysia, Kuala Lumpur.

—— (2000), World Tourism Organization, Annual Report of World Tourism Organization In Asia And The Pacific, 2000. http://www.world-tourism.org/regional/east_asia_and_pacific/pdf/wto_asia_pacific_2000.pdf. Accessed on 5 August 2002.