



International Journal of Emerging Technologies in Computational and Applied Sciences (IJETCAS)

www.iasir.net

Forecasting foreign tourist arrivals in India using different time series models

Kriti Kumari

Department of Mathematics and Statistics
Banasthali Vidyapith, Rajasthan, India

Abstract: In this study three time series models is used for forecasting monthly foreign tourist arrivals (FTAs) in India from January 2000 to October 2015. This paper compare the forecasting accuracy of seasonal autoregressive integrated moving average (SARIMA), Holt Winters (HW) and Grey model (GM) using mean absolute percentage error (MAPE) criterion. Moreover, turning point analysis (TPA), U statistics and posterior variance ratio test (PVRT) has also been carried out to strengthen decisions, and also to check accuracy of various forecasting models. Based on the forecasting performance of all three models, SARIMA and HW are found to be best models as compare to Grey model.

Keywords: Forecasting, tourist arrivals, SARIMA, Holt Winters model, Grey model

I. Introduction

Tourism industry is rising rapidly over the past few decades and it has huge prospective for producing employment and earning large amount of foreign exchange to the country's overall economic and social development. In developing countries like India, tourism has become one of the major sectors of the economy, contributing to a large proportion of the national income and generating vast employment opportunities. It has become the fastest growing service industry in the country with its great potentials for its further expansion and diversifications. Travelling and tourism has been an integral part of Indian culture and tradition. Tourism industry has also developed the significant source of employment, economic growth, foreign exchange earnings, revenue and income and also developing the infrastructure. Therefore, more accurate and reliable prediction of tourism demand is required for efficiently and effectively estimation of FTAs which in turn assures the future planning of infrastructure in tourism sector.

From the past several decades, study on tourism forecasting has attracted much of attention of the researchers across the world, for instance [17]-[18] studied the forecasting of Malaysian tourist arrivals and foreign tourist arrivals in Australia. [3]-[4] examined the tourism demand in Asia Pacific regions using various forecasting approaches. Recently, [19] compared the various forecasting models in predicting the tourism demand in Vietnam.

Since seasonality is an integral component of tourism demand, researchers found that among the competing time series models, SARIMA found to be more appropriate model for non-seasonal and seasonal time series forecasting [3, 5 and 15]. Additionally, SARIMA model is popular due to its ability to deal with both stationary and non-stationary series. However, [6] demonstrates that SARIMA model is the best one for forecasting inbound air travel arrivals to Taiwan. Among different time series forecasting models, Holt Winters has also gained more popularity for forecasting seasonal time series. Holt Winters is frequently adopted in empirical studies, because it often outperforms many time series models [see 3 and 6]. SARIMA and Holt Winters are seems to be more accurate models for seasonal tourist arrivals prediction.

In recent years, Grey model have been commonly used in tourism demand forecasting. For instance, [23] studied the prediction of tourism demand in Taiwan from the United States, Hong Kong and Germany using fuzzy time series and Grey model. However, [6] have been applied the Grey model to forecast inbound air travel arrivals to Taiwan. Grey model and its applications have been successfully conducted in tourist arrivals forecasting. Recently, [19 and 21] proposed Grey model to study the forecasting of foreign tourist arrivals in Vietnam and China.

To the best of our knowledge there is not much of literature available on the study of tourism demand in India. Though there are few, like [20] showed that SARIMA is the best model to forecast international tourists in India under MAPE criterion. In [7] the author developed a demand model of length of stay behaviour of international arrivals in India using poisson regression model and negative binomial analysis. There is a lot of scope to study the forecasting of tourism demand in India using different types of forecasting models. On this note, we suggest

to study the performance of SARIMA, Holt Winters and Grey model under different comparison criteria to forecast foreign tourist arrivals in India in the coming years.

This study is organised as follows. Section II describes the SARIMA, Holt Winters and Grey model. Section III presents the different performance criteria i.e. MAPE, TPA, U statistics and PVRT. Section IV describes the data. Section V provides the details of empirical results. Last section gives the summary and conclusions.

II. Forecasting models

A. SARIMA model

The SARIMA model was first developed by Box-Jenkins approach in [1]. It is used when the time series data contain seasonal elements. The SARIMA model has been created from both seasonal and non-seasonal autoregressive (AR), the moving average (MA) and the combination of AR and MA models (see Chu 1998 and Chen et al. 2009). It is usually represented by multiplicative model in the form of SARIMA (p,d,q)(P,D,Q)_s, where p, P indicates the non-seasonal order of AR and q, Q denote the seasonal order of MA model with non-seasonal difference (d) and seasonal difference (D). The best fitted SARIMA model is selected using minimum Akaike Information Criterion (AIC) [for details, see 2].

B. Holt Winters model

Tourism demand is generally supposed to be affected by trend and seasonal component. To confine these patterns, Holt [11] and Winters [22] integrate them into an exponential smoothing process. Holt Winters is an extension of the simple exponential and Holt smoothing model to allow the time series forecasting with trend and seasonal component [see 3 and 6]. There are two another approaches of Holt Winters model in seasonal time series forecasting. First is called additive Holt Winters (AHW), while the second is called multiplicative Holt Winters (MHW) [for details see 2].

C. Grey model

Grey forecasting is a single variable and first order model given by Deng [8], called as GM (1,1) model. GM (1,1) is one of the most frequently used in time series forecasting. This model can only be used in positive data sequences [10]. The Grey model adopts the essential part of the Grey system theory and it has been successfully used in the market for air travel [13]. The Grey model uses the operations of accumulated generation to build linear differential equation [see 21 and 23].

III. Measurements of forecasting accuracy

The performance of all such forecasting models for seasonal time series have been evaluated by different criteria like mean absolute percentage error (MAPE) [3 and 16], turning point analysis (TPA) [3 and 6] and U statistics [3]. To examine the accuracy of the SARIMA and HW models, three criteria i.e. MAPE, TPA and U statistics (U) are used. We have also applied the PVRT (C) and MAPE criteria for Grey model [see 19]. To check the forecasting accuracy of Grey model, four grades [9] are given below in Table 1.

Table 1: Grades for Grey forecasting model

Grades	C
Good	<0.35
Qualified	<0.5
Just	<0.65
Unqualified	≥0.65

IV. Data

In this study monthly foreign tourist arrivals (FTAs) from 2003:1-2013:12 in India is analysed. The FTAs data were obtained from <http://tourism.gov.in>. The FTAs series is divided into three periods (1) 2003:1-2012:12, consists of 120 observations, to fit the models; (2) data of 2012:1-12 (3) data of 2013:1-2015:10 are used to test the accuracy of in-sample and out-of-sample forecast.

V. Empirical findings

In this study, the forecasting performance of different models for FTAs series has been evaluated using R-3.0.3 version. Minitab-16 is also conducted for the time series plot.

A. Estimated SARIMA models

The time series plot (see Figure 1) shows that the FTAs series is non-stationary. The FTAs series are checked for unit root using the augmented Dickey–Fuller (ADF) test [see14]. The p-value (PV) of ADF test i.e. 0.4 with

trend and intercept is insignificant at 5% level of significance. Therefore, null hypothesis of unit root is not rejected, which indicates that FTAs series is non-stationary [see 17 and 18].

Since FTAs series existing seasonal patterns, the HEGY test procedure [see 12] has been developed to check the seasonal unit roots in the series. The results of HEGY test are presented in Table 2. From the Table 2, many p-values of seasonal dummy variables (SDV) and seasonal unit roots (SUR) are insignificant at 5% level of significance [see 17 and 18]. Hence, the null hypothesis of seasonal unit roots is not to be rejected at 5% level of significance and we conclude that the presence of deterministic as well as stochastic seasonality in the series [see 12]. In next step non-seasonal first difference ($d = 1$) is used to transformed non-stationary into stationary series. And seasonal twelve difference ($D = 12$) is conducted to remove seasonality from the series. Then the orders of p , q and P , Q for the SARIMA models are examined from autocorrelation (AC) and partial autocorrelation (PAC) functions [19]. The best fitted model is selected based on minimum Akaike Information Criterion (AIC). Table 3 describes the SARIMA models for the FTAs series. Further, from the Table 3 the best fitted model for FTAs series is $(2, 1, 1)(2, 1, 1)_{12}$ according to the minimum AIC criterion.

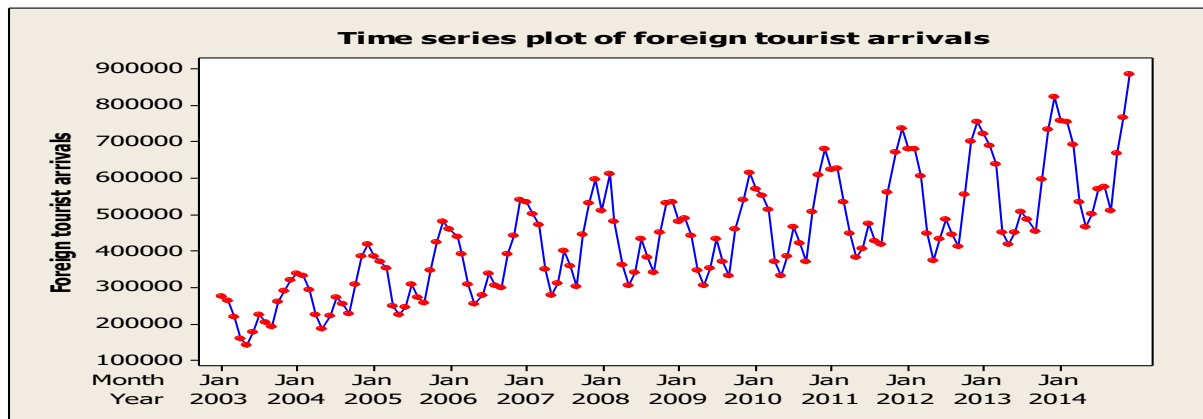


Figure 1: Time series plot of foreign tourist arrivals (in Numbers).

Table 2: Results of HEGY test

SDV	Coefficient	PV	SUR	Coefficient	PV	SUR	Coefficient	PV
C	18395.83	0.22	$Z_1(t-1)$	-0.001517	0.54	$Z_8(t-1)$	-0.045145	0.74
S_1	-12370.3	0.65	$Z_2(t-1)$	-0.117905	0.08	$Z_8(t-2)$	0.02867	0.83
S_2	-8637.103	0.75	$Z_3(t-1)$	-0.026087	0.66	$Z_8(t-3)$	0.254242	0.06
S_3	1896.525	0.94	$Z_3(t-2)$	-0.133436	0.02	$Z_8(t-4)$	-0.137239	0.32
S_4	-51944.28	0.06	$Z_4(t-1)$	-0.190249	0.09	$Z_8(t-5)$	0.016055	0.90
S_5	-21558.02	0.44	$Z_4(t-2)$	-0.287347	0.01	$Z_8(t-6)$	0.067756	0.62
S_6	7174.773	0.79	$Z_5(t-1)$	0.006268	0.77	$Z_8(t-7)$	-0.073616	0.59
S_7	42810.94	0.12	$Z_5(t-2)$	-0.014579	0.50	$Z_8(t-8)$	0.082209	0.54
S_8	-21495.88	0.43	$Z_6(t-1)$	-0.239468	0.02	$Z_8(t-9)$	0.068924	0.61
S_9	-80728.15	0.00	$Z_6(t-2)$	-0.293966	0.00	$Z_8(t-10)$	-0.081869	0.55
S_{10}	42295.23	0.14	$Z_7(t-1)$	-0.040847	0.52	$Z_8(t-11)$	-0.05746	0.71
S_{11}	40846.11	0.16	$Z_7(t-2)$	-0.076459	0.23	$Z_8(t-12)$	-0.029953	0.80

Table 3: Estimated SARIMA models

Models	AIC
$(2, 1, 1)(2, 1, 1)_{12}$	2236.90
$(2, 1, 1)(0, 1, 1)_{12}$	2247.36
$(1, 1, 1)(1, 1, 1)_{12}$	2241.67
$(0, 1, 1)(2, 1, 1)_{12}$	2246.54

B. Forecasting accuracy of Grey model using PVRT criterion

In this paper, we showed that GM (1,1) is not a good model to forecast seasonal FTAs series. A seasonal time series is to be deseasonalized to eliminate the seasonality from the series. Then, it is a better model to forecasting deseasonal FTAs series. The results of PVRT are shown in Table 4 for seasonal and deseasonal FTAs data. Comparing the value of C in Table 4 from the forecasting grades in Table 1, GM (1,1) is found to be unqualified model for forecasting seasonal tourist arrivals. Also, it is a good model for deseasonal tourist arrivals forecasting.

Table 4: Results of PVRT for Grey model

Series	C
Seasonal	0.70
Deseasonal	0.29

C. Comparison of Grey model with SARIMA and Holt Winters under MAPE criterion

Here, we present comparisons of two forms of Holt Winters model i.e. AHW and MHW models along with SARIMA and Grey model using MAPE criterion. The results obtained are reported below:

Table 5: MAPE of forecasting models for in-sample and out-of-sample (in %)

Sample	In-sample	Out-of-sample		
Models	2012	2013	2014	2015
SARIMA	3.44	2.98	3.50	3.70
AHW	5.10	5.21	4.24	6.44
MHW	2.54	2.54	4.28	4.01
GM	23.63	23.65	20.0	22.09

Table 5 reveals that the SARIMA and MHW has lesser MAPE (in-samples and out-samples) as compared to all other forecasting models. Between AHW and MHW, it is reasonable to choose MHW for further decision making process. Also, the forecasting performance of Grey model showed to be worse among all under MAPE criterion. Hence, following [16] SARIMA and MHW are the best models for in-sample and out-sample forecasts. Therefore, in the next section, U statistic and TPA has been performed for SARIMA and MHW models.

D. Analysis of TPA and U statistic based on SARIMA and MHW models

In this research TPA and U statistic has been also explored to evaluate the performance of SARIMA and MHW models. In Table 6 the positive and negative signs indicate that the positive and negative changes in actual and forecasted arrivals. If the actual changes (AC) and forecasted changes (FC) has different signs then turning point errors occur [3]. Graphical representation of actual and forecasted changes is presented in Figure 2-3 for both the models.

Table 6: Actual and forecasted changes based on SARIMA and MHW models

Months/Year	SARIMA				MHW		
	Actual FTAs	Forecasted FTAs	AC	FC	Forecasted FTAs	AC	FC
2013							
Jan	720321	697695			69193		
Feb	688569	694143	-	-	69948	-	+
Mar	639530	623821	-	-	62139	-	-
Apr	450580	486160	-	-	47553	-	-
May	417453	424088	-	-	40352	-	-
Jun	451223	472829	+	+	45208	+	+
Jul	506427	537453	+	+	53714	+	+
Aug	486338	493394	-	-	48983	-	-
Sep	453561	459018	-	-	45173	-	-
Oct	598095	600165	+	+	61347	+	+
Nov	733923	723500	+	+	73877	+	+
Dec	821581	783683	+	+	79958	+	+

Figure 2-3 shows the closeness of forecasted and actual changes. All scatter points are very close to the diagonal line of perfect forecast for SARIMA and MHW models. Table 7 shows the TPE and U statistics for the two good forecasting models. The values of U for both the models are very close to 0 [see 3] (out-of-sample), it indicates good accuracy of forecasting. Overall, SARIMA and MHW are highly accurate models for the future prediction of seasonal FTAs. Hence, the actual, fitted and long term future flows from 2015:10-2020:12 of FTAs series using SARIMA and MHW models are shown in Figure 4-5.

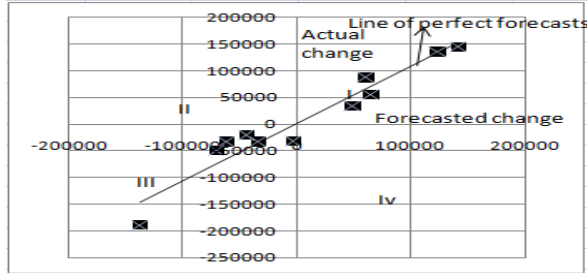


Figure 2: AC and FC of SARIMA model.

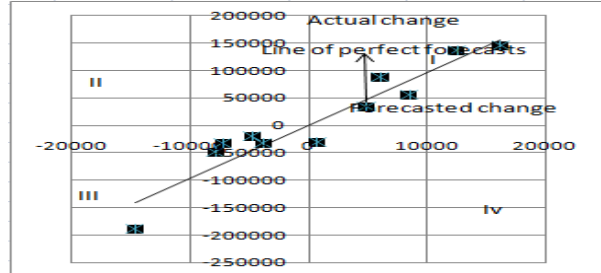


Figure 3: AC and FC of MHW model.

Table 7: TPE and U statistics of SARIMA and MHW models for out-of-sample

Criteria		U			TPE		
Years	2013	2014	2015	2013	2014	2015	
Models							
SARIMA	0.017	0.024	0.020	0	1	0	
MHW	0.014	0.022	0.023	1	2	1	

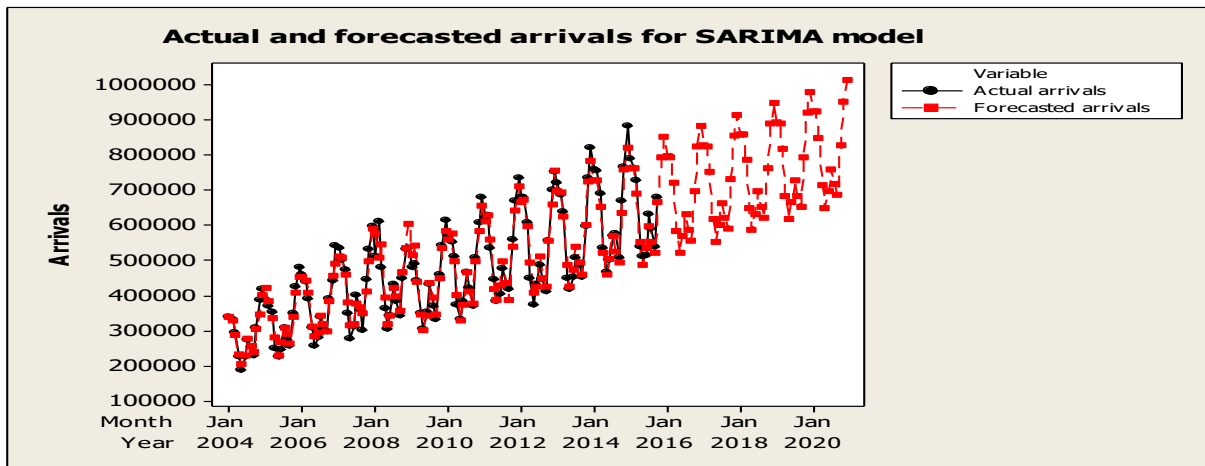


Figure 4: Actual and forecasted arrivals for SARIMA model.

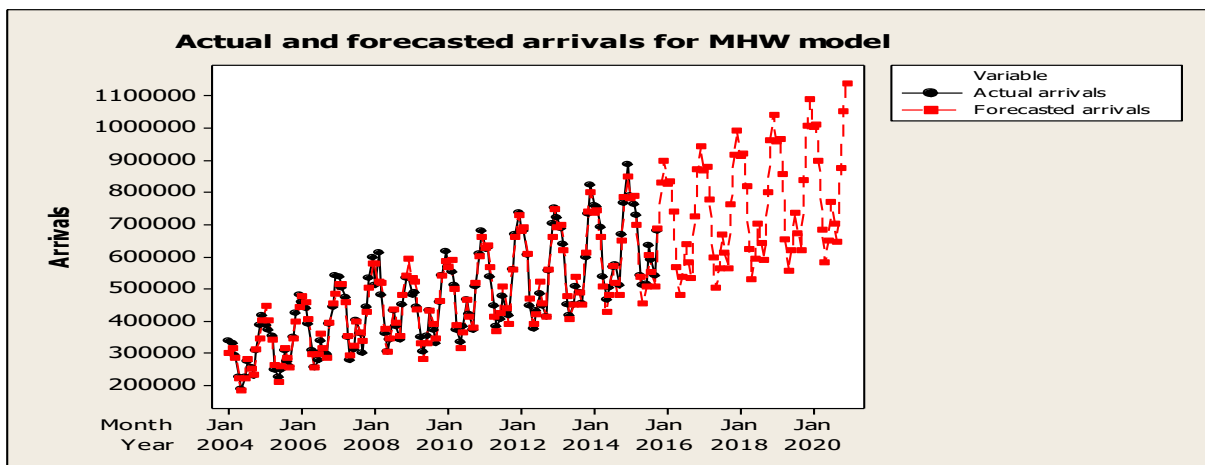


Figure 5: Actual and forecasted arrivals for MHW model.

VI. Conclusions

A perfect and accurate tourism demand prediction plays a vital role in the making of successful development policy for the government. The current and future trend of tourist arrivals specifies the future of international tourism demand. Therefore this research proposed a long term future demand of foreign tourist arrivals in India. In this study we evaluated the forecasting performance of SARIMA, Holt Winters and Grey model for FTAs series in India from 2003:1-2015:10 using MAPE criterion. Turning point analysis and U statistic has been also conducted for SARIMA and MHW models. In addition to, we carried out the posterior variance ratio test to check the forecasting accuracy of Grey model. Empirical findings show that the SARIMA and MHW models outperform the other models when compared under the MAPE criterion. The forecasting performance of Grey model proved to be worse among all under MAPE and PVRT criteria. Although, it is observed that the Grey model is a better option for a deseasonal series rather than seasonal series. However, turning point analysis and U statistic show that SARIMA and MHW are good models with highly forecasted accuracy.

References

- [1] Box, G. E. P. and Jenkins, G. M. (1976): "Time series analysis", Forecasting and Control ".
- [2] Chatfield, C. (1995): "The analysis of time series an introduction", Chapman and Hall.J.
- [3] Chu, F. L. (1998): "Forecasting tourism demand in Asian Pacific countries", Annual of Tourism Research, 25(3), 597-615.
- [4] Chu, F. L. (2004): "Forecasting tourism demand: a cubic polynomial approach", Tourism Management, 25(2), 209-218.
- [5] Coshall, J. (2006): "Time series analysis of UK outbound travel by air", Journal of Travel Research, 44(3), 335-347.
- [6] Chen, C. F. and Chang, Y.H. and Chang, Y.W. (2009): "Seasonal ARIMA forecasting of inbound air travel arrivals to Taiwan ", Transportmetrica, 5(2), 125-140.
- [7] Chaiboonsri, C. and Chaitip, P. (2012): " The modeling International tourism demand for length of stay in India: Social Development, Economics Development", Journal of Knowledge Management, Economics and Information Technology, 2, 1-17.
- [8] Deng, J. (1982): "Control problems of Grey systems, "Systems and Control Letters, 5(2), 288-294.
- [9] Deng, J. (1986): "Grey prediction and decision. Huazhong University of Science and Technology, Wuhan, China (in Chinese).
- [10] Deng, J. (1989): "Introduction to Grey system theory", The Journal of Grey system, 1, 1-24.
- [11] Holt, C. C. (1957): "Forecasting seasonals and trends by exponentially weighted moving averages. Pittsburgh PA: Carnegie Institute of Technology.
- [12] Hylleberg, S., Engle, R.F., Granger, C.W.J. and Yoo, B.S. (1990): "Seasonal integration and cointegration", Journal of Econometrics, 44 (2), 215-238.
- [13] Hsu, C. I. and Wen, Y. U. (1998). Improved Grey prediction models for trans-pacific air passenger market. Transportation Planning and Technology, 22, 87–107.
- [14] Lutkepohl, H. and Kratzig, M. (2004): "Applied time series econometrics", Cambridge University Press, New York.
- [15] Kulendran, N. and Wong, K. K. F. (2005): "Modelling seasonality in tourism forecasting", Journal of Travel Research, 44(2), 163-170.
- [16] Lewis, C.D. (1982): "International and business forecasting methods", London: Butterworths.
- [17] Lim, C. and McAleer, M. (1999): "A seasonal analysis of Malaysian tourist arrivals to Australia", Mathematics and Computers in Simulation, 48(6), 573-583.
- [18] Lim, C. and McAleer, M. (2002): "Time series forecasts of international travel demand for Australia", Tourism Management, 23(4), 389-396.
- [19] Nguyen, L.T., Shu, H.M., Huang, F.Y. and Hsu, M.B. (2013): " Accurate forecasting models in predicting the inbound tourism demand in Vietnam", Journal of Statistics and Management Systems, 16(1), 25-43.
- [20] Padhan, P.C. (2011): "Forecasting International Tourists Footfalls in India: An assortment of competing Models", International Journal of Business and Management, 6(5), 190-202.
- [21] Sun, X., Sun, W., Wang, J., Zhang, Y. and Gao, Y. (2016): " Using a Grey-Markov model optimized by Cuckoo search algorithm to forecast the annual foreign tourist arrivals to China", Tourism Management, 52, 369-379.
- [22] Winters, P.R., (1960): "Forecasting sales by exponentially weighted moving averages", Management Science, 6(3), 324-342.
- [23] Wang, C.H. (2004) "Predicting tourism demand using fuzzy time series and hybrid Grey theory", Tourism Management, 25(3), 367-374.