Forecasting with Fourier Residual Modified ARIMA Model- An Empirical Case of Inbound Tourism Demand in New Zealand

YING-FANG HUANG[1], PENG-JEN CHEN[1], THANH-LAM NGUYEN[2,*]
[1] Department of Industrial Engineering and Management
[2] Graduate Institute of Mechanical and Precision Engineering
National Kaohsiung University of Applied Sciences
415, Chien Kung Road, Sanmin District, Kaohsiung 80778
TAIWAN, R.O.C
(* Corresponding author: green4rest.vn@gmail.com)

Abstract: - Tourism, one of the biggest industries in many countries, has been considered a complexly integrated and self-contained economic activity. As key determinants of the tourism demand are not fully identified, to some extent different forecasting models vary in the level of accuracy. By comparing the performance of diverse forecasting models, including the autoregressive integrated moving average (ARIMA) and its joint Fourier modified model, this paper aims at obtaining an efficient model to forecast the tourism demand. The accuracy of the conventional model is found to be significantly boosted with the Fourier modification joined. In the empirical case study of inbound tourism demand in New Zealand, the Fourier modified seasonal ARIMA model FSARIMA(1,0,1)$(1,1,1)_{12}$ is strongly suggested due to its satisfactorily high forecasting power. We further employ the model to provide the New Zealand's tourism demand in 2013 so as to assist policy-makers and related organizations in establishing their appropriate strategies for sustainable growth of the industry.

Key-Words: - ARIMA model, Tourism demand, Fourier modification, New Zealand tourism, Tourism forecasting

1 Introduction

Nowadays, tourism has been considered as a "smokeless" and important industry in numerous countries in the world because it can not only generate plenty of quality jobs but also offer great contribution to the GDP. In New Zealand, the number of international tourist arrivals in 2012 was about 50% higher than that in 2000. Based on the annual research by the World travel & tourism council [1], in 2012, the total contribution of the tourism industry to New Zealand GDP was about NZD31.1 billion; accounting for 14.9% of GDP; and it supported 19.1% of the total employment with about 426.5 thousand jobs. In regarding to its direct contribution, the tourism contributed about NZD7.0 billion; accounting for 3.4% of total GDP and supported 133 thousand jobs (6% of total employment). These figures indicate that tourism is an important industry in New Zealand. In order to make proper plans for the sustainable development of the national tourism industry, accurately forecasting the tourism demand becomes critical.

In spite of its aforementioned importance, tourism has been considered not only as an integrated and self-contained economic activity without a strong support from economic theories but also as a complex system due to a strong inter-relationship existing among different dependable sectors in the economy such as economic, transportation, commerce, social & cultural services, political and technological changes, etc., [2]. As several determinants of the inbound tourism demand are neither easily measured nor collected due to their availability [2-6]. Also, as the number of tourist arrival has been widely used as an appropriate indicator of inbound tourism demand in many researches [3, 6-15], in this study, the monthly arrivals of inbound tourists to New Zealand from January 2000 to March 2013 are therefore used to denote the inbound tourism demand in New Zealand.

As tourism is season-sensitive with the inherent characteristic of a time series, it is therefore suggested to use autoregressive integrated moving average (ARIMA), a well-known forecasting model dealing with time series, to predict the demand. The residual series obtained from this traditional model is then modified with Fourier series so as to improve the model accuracy. In order to evaluate the forecasting power of the Fourier modified model, we compare the forecast values with the actual ones from January 2013 – March 2013 before being further employed to have a longer forecast.

2 Literature Review

2.1 ARIMA Model

ARIMA model was first introduced by Box and Jenkins in 1960s to forecast a time series which can be made stationary by differencing or logging. A time series may have non-seasonal or seasonal characteristics. Seasonality in a time series is defined as a regular pattern of changes that repeats over S time-periods. With a seasonal time series, there is usually a difference between the average values at some particular times within the seasonal intervals and the average values at other times; therefore, in most cases, a seasonal time series is non-stationary.

2.1.1 Non-seasonal ARIMA model

A non-seasonal ARIMA model usually has the form of *ARIMA(p,d,q)*, where:

- p is the number of lags of the differenced series appeared in the forecasting equation, called auto-regressive parameter,
- d is the difference levels to make a time series stationary, called integrated parameter, and
- q is the number of the lags of the forecast errors, called moving-average parameter.

2.1.2 Seasonal ARIMA model

With a seasonal time series, it can be made stationary by seasonal differencing which is defined as a difference between a value and a value with lag that is a multiple of S.

The Seasonal ARIMA model incorporates both non-seasonal and seasonal factors in a multiplicative model with the form of *SARIMA(p,d,q)(P,D,Q)$_S$,* where:

- p, d, q are the parameters in non-seasonal ARIMA model as mentioned above.
- P is the number of seasonal Autoregressive order,
- D is the number of seasonal differencing,
- Q is the number of seasonal Moving Average order, and
- S is the time span of repeating seasonal pattern.

There are three basic steps in the overall procedures to obtain an ARIMA or SARIMA model [16], including:

**** Step 1: Identifying the possible models***

- Examine the Auto-Correlation Function (ACF) and Partial Auto-Correlation Function (PACF) graphs to identify non-seasonal terms.

The time series is considered stationary if the lag values of the ACF cut off or die down fairly quickly. If the series is not stationary, it should be differenced gradually until it is considered stationary. Then, the d value in the model is obtained. If ACF graph cut off after lags q fairly quickly and PACF graph cut off after lags p fairly quickly, ARMA(p,q) is achieved. ARIMA(p,d,q) is accordingly identified.

- Examine the patterns across lags that are multiples of S to identify seasonal terms. Judge the ACF and PACF at the seasonal lags in the same way.

**** Step 2: Fitting the model***

In this step, the parameters of the model are estimated.

**** Step 3: Testing the model for adequacy***

The residuals from the model must have normal distribution and be white-noise (also known random). This test can be done with one of the following ways:

- Ljung-Box Q statistic [17]:

$$Q_m = n(n+2)\sum_{k=1}^{m}\frac{e_k^2}{n-k} \tag{1}$$

where: e_k is the residual autocorrelation at lag k; n is the number of residuals; and, m is the number of time lags includes in the test.

The model is considered adequate only if the p-value associated with the Ljung-Box Q Statistic is higher than a given significance.

- Considering the histogram plot of the residual series to make sure it has normal distribution and considering its ACF and PACF graphs where no peak is found.

2.2 Fourier Residual Modification

Grey forecasting models have been proved to be significantly improved after their residual series are modified with Fourier series [18-22]. So, this good methodology should also be considered in the case of ARIMA model. The procedure to obtain the modified model is as the following.

Based on the predicted series $\hat{x}^{(0)}$ obtained from the ARIMA model, a residual series named $\varepsilon^{(0)}$ is defined as:

$$\varepsilon^{(0)} = \left\{\varepsilon_2^{(0)}, \varepsilon_3^{(0)}, \ldots, \varepsilon_k^{(0)}, \ldots, \varepsilon_n^{(0)}\right\} \quad (2)$$

where $\varepsilon_k^{(0)} = x_k^{(0)} - \hat{x}_k^{(0)} \quad \left(k = \overline{2,n}\right)$

Expressed in Fourier series, $\varepsilon_k^{(0)}$ is rewritten as:

$$\varepsilon_k^{(0)} = \frac{1}{2}a_0 + \sum_{i=1}^{F}\left[a_i \cos\left(\frac{2\pi i}{n-1}k\right) + b_i \sin\left(\frac{2\pi i}{n-1}k\right)\right] \quad (3)$$

where $F = \left[(n-1)/2 - 1\right]$ called the minimum deployment frequency of Fourier series [22] and only take integer number [18, 19, 21]. And therefore, the residual series is rewritten as:

$$\varepsilon^{(0)} = P.C \quad (4)$$

where

$$P = \begin{bmatrix} \frac{1}{2} & \cos\left(\frac{2\pi\times1}{n-1}\times2\right) & \sin\left(\frac{2\pi\times1}{n-1}\times2\right) & \cdots & \cos\left(\frac{2\pi\times F}{n-1}\times2\right) & \sin\left(\frac{2\pi\times F}{n-1}\times2\right) \\ \frac{1}{2} & \cos\left(\frac{2\pi\times1}{n-1}\times3\right) & \sin\left(\frac{2\pi\times1}{n-1}\times3\right) & \cdots & \cos\left(\frac{2\pi\times F}{n-1}\times3\right) & \sin\left(\frac{2\pi\times F}{n-1}\times3\right) \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \frac{1}{2} & \cos\left(\frac{2\pi\times1}{n-1}\times n\right) & \sin\left(\frac{2\pi\times1}{n-1}\times n\right) & \cdots & \cos\left(\frac{2\pi\times F}{n-1}\times n\right) & \sin\left(\frac{2\pi\times F}{n-1}\times n\right) \end{bmatrix}$$

$$C = \left[a_0, a_1, b_1, a_2, b_2, \ldots, a_F, b_F\right]^T$$

The parameters $a_0, a_1, b_1, a_2, b_2, \ldots, a_F, b_F$ are obtained by using the ordinary least squares method (OLS) which results in the equation of:

$$C = \left(P^T P\right)^{-1} P^T \left[\varepsilon^{(0)}\right]^T \quad (5)$$

Once the parameters are calculated, the predicted series residual $\hat{\varepsilon}^{(0)}$ is then easily achieved based on the following expression:

$$\hat{\varepsilon}_k^{(0)} = \frac{1}{2}a_0 + \sum_{i=1}^{F}\left[a_i \cos\left(\frac{2\pi i}{n-1}k\right) + b_i \sin\left(\frac{2\pi i}{n-1}k\right)\right] \quad (6)$$

Therefore, based the predicted series $\hat{x}^{(0)}$ obtained from ARIMA model, the predicted series $\tilde{x}^{(0)}$ of the modified model is determined by:

$$\tilde{x}^{(0)} = \left\{\tilde{x}_1^{(0)}, \tilde{x}_2^{(0)}, \ldots, \tilde{x}_k^{(0)}, \ldots, \tilde{x}_n^{(0)}\right\} \quad (7)$$

where

$$\begin{cases} \tilde{x}_1^{(0)} = \hat{x}_1^{(0)} \\ \tilde{x}_k^{(0)} = \hat{x}_k^{(0)} + \hat{\varepsilon}_k^{(0)} \quad \left(k = \overline{2,n}\right) \end{cases}$$

In order to evaluate the accuracy of the forecasting model, the residual error (ε) and its relative error ($\wp$) are used [19, 23]. ε and $\wp$ of an entry k are expressed as:

- Residual error: $\varepsilon_k = x_k^{(0)} - f_k^{(0)} \quad \left(k = \overline{1,n}\right)$

  where $f_k^{(0)}$ is the forecasted value at the k^{th} entry
- Relative error: $\wp_k = \left|\varepsilon_k\right| / x_k^{(0)} \quad \left(k = \overline{1,n}\right)$

However, there have been some other important indexes to be considered in evaluating the model accuracy. They are:

- The mean absolute percentage error (MAPE) [5, 19-22, 24-26]:

$$MAPE = \frac{1}{n}\sum_{k=1}^{n}\wp_k$$

- The post-error ratio C [27, 28]:

$$C = \frac{S_2}{S_1}$$

where: $S_1 = \sqrt{\frac{1}{n}\sum_{k=1}^{n}\left[x_k^{(0)} - \frac{1}{n}\sum_{k=1}^{n}x_k^{(0)}\right]^2}$

$$S_2 = \sqrt{\frac{1}{n}\sum_{k=1}^{n}\left[\varepsilon_k - \frac{1}{n}\sum_{k=1}^{n}\varepsilon_k\right]^2}$$

The ratio C, in fact, is the ratio between the standard deviation of the original series and the standard deviation of the forecasting error. The smaller the C value, the higher accuracy the model has since smaller C value results from a larger S_1 and/or a smaller S_2.

- The small error probability P [27, 28]:

$$P = p\left\{\left(\left|\varepsilon_k - \frac{1}{n}\sum_{k=1}^{n}\varepsilon_k\right| / S_1\right) < 0.6745\right\}$$

The higher the P value is, the higher accuracy the model has since P value indicates the probability of the ratio of the difference between the residual values of data points and the average residual value with the standard deviation of the original series smaller than 0.6745 [28].

- The forecasting accuracy ρ [28]:

$$\rho = 1 - MAPE\ .$$

Based on the above indexes, there are four grades of accuracy as stated in Table 1.

Table 1: Four grades of forecasting accuracy

Grade level	$MAPE$	C	P	ρ
I (Very good)	< 0.01	< 0.35	> 0.95	> 0.95
II (Good)	< 0.05	< 0.50	> 0.80	> 0.90
III (Qualified)	< 0.10	< 0.65	> 0.70	> 0.85
IV (Unqualified)	≥ 0.10	≥ 0.65	≤ 0.70	≤ 0.85

3 Empirical Study

Historical data of the inbound tourism demand in New Zealand from January 2000 – March 2013 (totally 160 observations) are obtained from the monthly statistical data published by Statistics New Zealand [29]. The data from January 2000 – December 2012 are used to build an ARIMA/ SARIMA model which is then modified with Fourier series to become a modified model with higher accuracy. Data from January 2013 – March 2013 are used to check the forecast power of the modified model before it is employed to forecast the demand in the other three quarters of 2013.

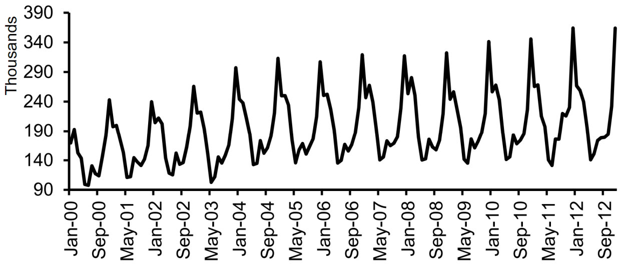

Figure 1. Monthly inbound arrivals to New Zealand

From Figure 1, it can be concluded that seasonality exists in the series of tourism demand. Therefore, only seasonal ARIMA model is considered in this section. At one degree of seasonal difference, the series becomes stationary as shown in Figure 2 and Figure 3.

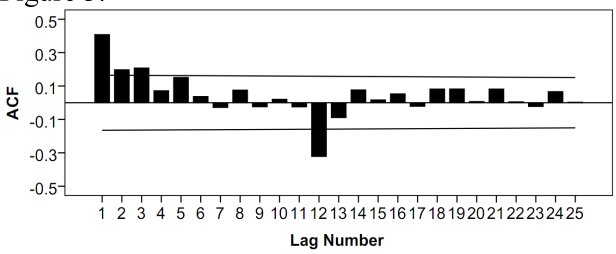

Figure 2. Auto-correlation function graph

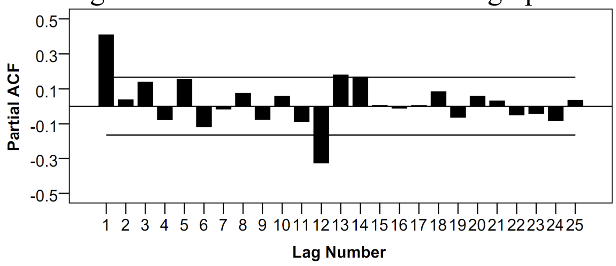

Figure 3. Partial Auto-correlation function graph

From the Figure 2 and Figure 3, there are three possible SARIMA models as the following.

- Model 1: SARIMA(1,0,1)(1,1,1)$_{12}$
- Model 2: SARIMA(1,0,2)(1,1,1)$_{12}$
- Model 3: SARIMA(1,0,3)(1,1,1)$_{12}$

illustrated in Table 2.

Table 2. Model summary statistics

Model		Model 1	Model 2	Model 3
R-Squared		0.963	0.963	0.963
MAPE		4.266	4.361	4.376
MAE		8034.843	8143.048	8167.646
Ljung -Box	Stat.	10.365	11.099	10.909
	Df	14	13	12
	Sig.	0.735	0.603	0.537

Based on Table 2, among the three models, Model 1 is the best. Figure 4 shows that Model 1 is adequate; it is, therefore, selected. The residual series obtained from the selected model SARIMA(1,0,1)(1,1,1)$_{12}$ is now modified with Fourier series, making the model become a new one- FSARIMA(1,0,1)(1,1,1)$_{12}$. The evaluation of these two models is shown in Table 3.

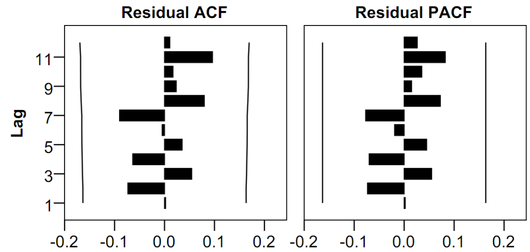

Figure 4. Noise residual ACF-PACF

Table 3: Evaluation indexes of model accuracy

Index \ Model	SARIMA	FSARIMA
MAPE	0.0427	0.0069
S_1	55552.12	55552.12
S_2	10690.32	3029.26
C	0.1924	0.0545
P	0.9931	1.0000
ρ	0.9573	0.9931
Forecasting power	Good	Very good

In order to further evaluate the forecasting power of FSARIMA(1,0,1)(1,1,1)$_{12}$, we now compare the forecast value in January-March 2013 with the actual observations in the same period, which results in Table 4. With the MAPE value of 0.0202, FSARIMA(1,0,1)(1,1,1)$_{12}$ is considered powerful to be employed to forecast the number of inbound arrivals from April – December 2013 as shown in Table 5.

Table 4. Checking forecasting power

Month	Actual	Forecast	APE
Jan. 2013	260,637	262,918	0.0183
Feb. 2013	281,233	273,325	0.0281
Mar. 2013	270,740	266,862	0.0143
Mean absolute percentage error			0.0202

Table 5. Forecast in 2013 (Unit: Arrivals)

Month	Forecast	Month	Forecast
Apr. 2013	264730	Sept. 2013	268427
May. 2013	209433	Oct. 2013	269342
Jun. 2013	209480	Nov. 2013	299968
Jul. 2013	244215	Dec. 2013	372126
Aug. 2013	245428		

4 Conclusion

The accuracy level of a traditional ARIMA model can be significantly improved with the procedure called Fourier residual modification. Particularly, in the case of the inbound tourism demand in New Zealand, despite of the hard assessment of relevant data about its determinants, the monthly demand can still be forecasted effectively with a modified model FSARIMA(1,0,1)(1,1,1)$_{12}$. Precise forecasting result will help the policy-makers and related organizations in the tourism industry to arrange enough facilities and human resources for high seasons and also make regular maintenance and training in low seasons just for a stable growth of the industry.

References:

[1] World Travel and Tourism Council, 2013 Annual Research: Key Facts. Available from: *http://www.wttc.org/research/economic-impact-research/country-reports/n/New Zealand/*

[2] Lagos, D.G, Exploratory forecasting methodologies for tourism demand, *Ekistics Journal*, Vol. 396, 1999, pp. 143-154.

[3] González, P. and Moral, P., An analysis of the international tourism demand in Spain, *International Journal of Forecasting*, Vol. 11, 1995, pp. 233-251.

[4] Witt, S.F., and Witt, C.A., Forecasting tourism demand: A review of empirical research, *International Journal of Forecasting*, Vol. 11, 1995, pp. 447-475.

[5] Hsu, L.C., and Wang, C.H., The development and testing of a modified Diffusion model for predicting tourism demand, *International Journal of Management*, Vol. 25, 2008, pp. 439-445.

[6] Habibi, F., Rahim, K.A., Ramchandran, S., and Chin, L., Dynamic Model for International Tourism Demand for Malaysia: Panel Data Evidence, *International Research Journal of Finance and Economics*, Vol. 33, 2009, pp. 207-217.

[7] Ouerfelli, C., Co-integration analysis of quarterly European tourism demand in Tunisia, *Tourism Management*, Vol. 29, 2008, pp. 127-137.

[8] Morley, C.L., A dynamic international demand model, *Annals of Tourism Research*, Vol. 23, 1998, pp. 70-84.

[9] Lim, C., and McAleer, M., Monthly seasonal variations: Asian tourism to Australia, *Annals of Tourism Research*, Vol. 28, 2001, pp. 68–82.

[10] Kulendran, N., and Witt, S., Co-integration versus least squares regression, *Annals of Tourism Research*, Vol. 28, 2001, pp. 291–311.

[11] Tan, Y. F., McCahon, C., and Miller, J., Modelling tourist flows to Indonesia and Malaysia, *Journal of Travel and Tourism Marketing*, Vol. 12, 2002, pp. 63–84.

[12] Song, H. and Witt, S.F., General-to-specific modeling to international tourism demand forecasting, *Journal of Travel Research*, Vol. 42, 2003, pp. 65-74.

[13] Song, H., Witt, S. F., and Li, G., Modelling and forecasting the demand for Thai tourism, *Tourism Economics*, Vol. 9, 2003, pp. 363–387.

[14] Dritsakis, N., Co-integration analysis of German and British tourism demand for Greece, *Tourism Management*, Vol. 25, 2004, pp. 111-119.

[15] Naude, A.W., and Saayaman, A., Determinants of tourist arrivals in Africa: a panel data regression analysis, *Tourism Economics*, Vol. 11, 2005, pp. 365-391.

[16] Hanke, J.E. and Wichern, D.W., *Business Forecasting*. 8th Ed. Pearson, Prentice Hall, New Jersey, 2005.

[17] Ljung, G.M., and Box, G.E.P., On a measure of lack of fit in time series models, *Biometrika*, 65, 1978, 297-303.

[18] Hsu, L.C., Applying the grey prediction model to the global integrated circuit industry, *Technological forecasting & Social change*, Vol. 70, 2003, pp. 563-574.

[19] Guo, Z., Song, X. and Ye, J., A Verhulst model on time series error corrected for port throughput forecasting, *Journal of the Eastern Asia society for Transportation studies*, Vol. 6, 2005, pp. 881-891.

[20] Kan, M.L., Lee, Y.B. and Chen, W.C., Apply grey prediction in the number of Tourist, in *The fourth international conference on Genetic and Evolutionary computing in Shenzhen*, China, 2010, pp. 481-484.

[21] Askari, M., and Fetanat, A., Long-term load forecasting in power system: Grey system prediction-based models, *Journal of Applied Sciences*, Vol. 11, 2011, pp. 3034-3038.

[22] Huang, Y.L. and Lee, Y.H, Accurately forecasting model for the Stochastic Volatility data in tourism demand, *Modern economy*, Vol. 2, 2011, pp. 823-829.

[23] Liang, P., Zhou, H., and Zheng, J., Forecast research of droplet size based on Grey theory, *IFIP Advances in Information and Communication Technology*, Vol. 258, 2008, pp. 653-658.

[24] Hsu, C.C., and Chen, C.Y., Applications of improved grey prediction model for power demand forecasting, *Energy Conversion and Management*, Vol. 44, 2003, pp. 2241-2249.

[25] Chang, Y.W. and Liao, M.Y., A seasonal ARIMA model of tourism forecasting: The case of Taiwan, *Asia Pacific journal of Tourism research*, Vol. 15, 2010, pp. 215-221.

[26] Tsaur, R.C., and Kuo, T.C, The adaptive fuzzy time series model with an application to Taiwan's tourism demand, *Expert Systems with Applications*, Vol. 38, 2011, pp. 9164-9171.

[27] Hua, J., and Liang, Q., Application research of the Grey forecast in the Logistics demand forecast, *First International Workshop on Education Technology and Computer Science*, Vol. 3, 2009, pp. 361-363.

[28] Ma, H., and Zhang, Z., Grey prediction with Markov-Chain for Crude oil production and consumption in China, *Advances in Intelligent and Soft Computing*, Vol. 56, 2009, pp. 551-561.

[29] Statistics New Zealand (2013), *http://www.stats.govt.nz/infoshare/*