International Journal of Innovative Computing, Information and Control
Volume **10**, Number **2**, April **2014**

 ISSN 1349-4198
pp. 695–701

TOURISM DEMAND FORECASTING USING A NOVEL HIGH-PRECISION FUZZY TIME SERIES MODEL

Ruey-Chyn Tsaur and Ting-Chun Kuo

Department of Management Sciences
Tamkang University
No. 151, Yingzhuan Rd., Danshui Dist., New Taipei 25137, Taiwan
rctsaur@yahoo.com.tw

Received February 2013; revised June 2013

Abstract. *Fuzzy time series model has been developed to either improve forecasting accuracy or reduce computation time, whereas a residul analysis in order to improve its forecasting performance is still lack of consideration. In this paper, we propose a novel Fourier method to revise the analysis of residual terms, and then we illustrate it to forecast the Japanese tourists visiting in Taiwan per year. The forecasting results show that our proposed method can derive the best forecasting performance as well as the smallest forecasting error of MAPE in the training sets; in the testing sets, the proposed model is also better to fit the future trend than some forecasting models.*
Keywords: Fuzzy time series, Fourier series, Residual analysis, Japanese tourists, Forecasting performance

1. **Introduction.** In a business environment, forecasting is important for fitting the future trends. Especially, statistical models play a big role in forecasting because of its ease of applications. The topic of tourism demand forecasting focuses on traditional time series, including Qu and Lam [1] performed an OLS multiple regression model to identify the best exogenous variables to determine the demand model; Lim and McAleer [2] used ARIMA model to forecast the tourist to Australia from Hong Kong, Malaysia and Singapore; Song and Witt [3] used the VAR model to suggest that Macau is necessary to consider the increasing tourism demand from mainland China; Lim *et al.* [4] used the ARIMA model to investigate the dynamic relationship between tourism demand and real income of Japan; Cheng [5] examined the determinants of Hong Kong tourism demand for the top-three major tourist arrival countries with an error correction model. The above methods have their specified advantages, but it is still difficult to use the above models for tourism demand analysis when the collected data are limited, or linguistic time series data. Fuzzy time series model [6,7] has been developed for forecasting analysis. For example, Lee *et al.* [8] uesed their proposed fuzzy time series model for temperature prediction and TAIFEX forecasting; Kuo *et al.* [9] presented a hybrid forecasting method to solve the TAIFEX forecasting problem based on fuzzy time series and particle swarm optimization; Cheng *et al.* [10] presented an adaptive expectation model to forecast TAIFEX; Wong *et al.* [11] applied fuzzy time series models to forecast the amount of Taiwan export; Wang [12], and Wang and Hsu [13] proposed methods to forecast the tourism demand. Fuzzy time series is a well-applied forecasting method, whereas a residul analysis for enhancing its forecasting performance is still lack of considerations. In this study, the Fourier series method is used to transform the residuals into frequency spectra, select the low-frquency terms, and then derive better forecasting performance. Finally,

an illustration for forecasting the tourists from Japan is used to perform the advatage of the proposed model.

2. **Fuzzy Time Series Model with Fourier Method.** Fuzzy time series model used fuzzy logic method to forecast without considering the data type usually reduced its forecasting performance. Therefore, any adjusted method for these large forecasting error terms to be lower error terms should be considered to adjust forecasting performance. The idea of Fourier series is selected to transform the residuals into frequency spectra, select the low-frequency terms, filter out high-frequency terms which are supported to be noise, and then have well forecasting performance [14]. Let U be the universe of discourse, $U = \{u_1, u_2, \ldots, u_n\}$, and A_i be a fuzzy set in the U defined as Equation (1).

$$A_i = \mu_{Ai}(u_1)/u_1 + \mu_{Ai}(u_2)/u_2 + \ldots + \mu_{Ai}(u_n)/u_n \tag{1}$$

where μ_{Ai} is the membership function of the fuzzy set A_i, and $\mu_{Ai}(u_j)$ represents the grade of membership of u_j in A_i where $\mu_{Ai}(u_j) \in [0, 1]$ and $1 \leq i, j \leq n$. Then, a fuzzy time series model with Fourier method can be defined as follows [6,7].

Definition 2.1. *If $Y(t)$ $(t = 1, 2, \ldots, n)$ is a subset of R^1 in which the universe of fuzzy sets $f_i(t)$ $(i = 1, 2, \ldots, m)$ are defined and let $F(t)$ be collection of $f_i(t)$ $(i = 1, 2, \ldots, m)$, then $F(t)$ is called a fuzzy time series on $Y(t)$ $(t = 1, 2, \ldots, n)$.*

Definition 2.2. *Let $F(t)$ be a fuzzy time series caused by $F(t-1), F(t-2), \ldots,$ and $F(t-n)$, then this fuzzy logical relationship is defined as $F(t-1), F(t-2), \ldots, F(t-\lambda) \rightarrow F(t)$, and it is called the λ^{th} order fuzzy time series forecasting model.*

The procedure of modelling fuzzy time series model is described as follows.

Step 1. Define the universe of discourse U. We choose the minimum value $D_{\min}$ and the maximum value $D_{\max}$ in the historical data, then $U = [D_{\min} - D_1, D_{\max} + D_2]$, where D_1 and D_2 are two proper positive numbers.

Step 2. Define the fuzzy sets A_i on universe of discourse U. Partition the universe of discourse of the collected data into equal length of intervals $u_1, u_2, \ldots, u_n$, then the fuzzy set $A_i = \mu_{Ai}(u_1)/u_1 + \mu_{Ai}(u_2)/u_2 + \ldots + \mu_{Ai}(u_n)/u_n$, $\forall i = 1, 2, \ldots, n$, where $\mu_{Ai}(u_j)$ represents the grade of membership of u_j in A_i, $1 \leq i, j \leq n$.

Step 3. Establish the fuzzy logical relationships. Fuzzyfied all the collected data and use all the fuzzy sets to define the second-order fuzzy logical relationship as A_i, $A_{j \rightarrow} A_k$ where time $t-2$ and $t-1$ are in states i and j, and time t is in state k [15]. By the second-order fuzzy logical relationships, we can categorize the fuzzy logical relationship groups.

Step 4. Calculate the forecasted outputs. For any chosen second-order fuzzy logical relationship [1], if $F(t-2) = A_i$ and $F(t-1) = A_j$, then the forecasting $F(t)$ is derived by the following rules, where *Rule 1* and *Rule 2* are used in-sample prediction, and *Rule 3* is used for out-of-sample extrapolation.

Rule 1: If fuzzy logical relationship group is A_i, $A_{j \rightarrow} A_y$, then forecasting value $F(t) = m_y$, where m_y is the midpoint of u_y, $y = 1, 2, \ldots, n$.

Rule 2: If the fuzzy logical relationship groups are $A_i, A_{j \rightarrow} A_1, A_i, A_{j \rightarrow} A_2, \ldots, A_i, A_{j \rightarrow} A_p$, then the forecasting value $F(t) = (m_1 + m_2 + \ldots + m_p)/p$, where m_i is the midpoint of u_i, $i = 1, 2, \ldots, p$.

Rule 3: If the fuzzy logical relationship group is A_i, $A_{j \rightarrow} \#$, where "#" is the next unknown state, then the forecasting value $F(t) = (m_j \times W + m_i)/[W + (\lambda - 1)]$.

Here λ is a positive integer to represent the order of the fuzzy logical relationship, m_i and m_j, $1 \leq i, j \leq n$, are the midpoints of the intervals with respect to the latest two past linguistic values which are fitted to fuzzy sets A_i, A_j, $1 \leq i, j \leq n$, and W is the weight for the latest state A_j, $1 \leq i, j \leq n$ that is defined by decision maker. We defined the second-order fuzzy logical relationship $\lambda = 2$, then Rule 3 can be rewritten as Equation (2). If $[W/(W+1)]$ is set as smoothing constant θ, then Equation (5) can be revised as an exponential smoothing equation as Equation (3).

$$F(t) = [1/(W+1)] \times m_i + [W/(W+1)] \times m_j \tag{2}$$

$$F(t) = (1-\theta) \times m_i + \theta \times m_j, \quad 0 < \theta \leq 1 \tag{3}$$

Because W is decided by a decision maker, we suggest that if a decision maker is sure of the trend of the future from the historical data, then θ can be assigned a larger value; if not, θ should be assigned a smaller value.

Step 5. Adjustment using Fourier series. Theoretically, if the collected time series data is limited, then each data is equally important for forecasting analysis. In traditional approaches, a fitted forecasting model is based on the assumption of realizing the structure of the system and assumption of normal distribution; however, because of limited information, the normal distribution is usually violated, and only part of system structure could be fully captured. Therefore, residual analysis becomes quite important in order to reuse some possible useful information for revision the original forecasted value. The idea of Fourier series is used to transform the residuals into frequency spectra and then select low-frequency terms. On the other hand, Fourier approaches can filter out high-frequency terms, which are supposed to be noise, and then have very nice performance.

First, the residual series is derived as follows.

$$\varepsilon(k) = Y(k) - F(k), \quad k = 1, 2, \ldots, N \tag{4}$$

$$\boldsymbol{\varepsilon(k)} = [\varepsilon(1), \varepsilon(2), \ldots, \varepsilon(N)]^T, \tag{5}$$

where $Y(k)$ and $F(k)$ are the actual value and fuzzy time series forecasted value at time k, respectively, and the matrix of the residual series is defined as $\boldsymbol{\varepsilon(k)}$.

Second, Fourier series is used to catch the residual series with periodic phenomenon as shown in Equation (6). Then, the estimated residual series by Fourier series are derived as

$$\hat{e}(k) = \frac{1}{2}a_0 + \sum_{i=1}^{z}\left[a_i \cos\left(\frac{2\pi i}{T}k\right) + b_i \sin\left(\frac{2\pi i}{T}k\right)\right], \quad k = 1, 2, \ldots, n, \tag{6}$$

where T is the length of the residual series which is equal to n, and z is the minimum deployment frequency of Fourier series which is the integer portion of $[(n-1)/2]$. Then, the parameters a_0, a_i and b_i for $i = 1, 2, \ldots, z$ are estimated by least square method as

$$\boldsymbol{C} = (\boldsymbol{P^T P})^{-1}\boldsymbol{\varepsilon(k)}, \tag{7}$$

$$\boldsymbol{P} = \begin{bmatrix} \frac{1}{2} & \cos\left(\frac{2\pi}{T}\right) & \sin\left(\frac{2\pi}{T}\right) & \ldots & \ldots & \ldots & \cos\left(\frac{2\pi\, z}{T}\right) & \sin\left(\frac{2\pi\, z}{T}\right) \\ \frac{1}{2} & \cos\left(2\frac{2\pi}{T}\right) & \sin\left(2\frac{2\pi}{T}\right) & \ldots & \ldots & \ldots & \cos\left(2\frac{2\pi\, z}{T}\right) & \sin\left(2\frac{2\pi\, z}{T}\right) \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ \frac{1}{2} & \cos\left(N\frac{2\pi}{T}\right) & \sin\left(N\frac{2\pi}{T}\right) & \ldots & \ldots & \ldots & \cos\left(N\frac{2\pi\, z}{T}\right) & \sin\left(N\frac{2\pi\, z}{T}\right) \end{bmatrix} \tag{8}$$

where $C = [a_0, a_1, b_1, b_2, \ldots, a_z, b_z]^T$. Hence, the estimated residual series is obtained as

$$\boldsymbol{\hat{e}(k)} = [\hat{e}(1), \hat{e}(2), \ldots, \hat{e}(N), \hat{e}(N+1)]^T \tag{9}$$

$$\hat{Y}'(t+1) = F(t+1) + \hat{e}(t+1), \quad \forall t = 1, 2, \ldots, n, \tag{10}$$

where $F(t+1)$ and $\hat{e}(t+1)$ are forecasting value and estimated residual value at time $t+1$.

An illustration adopted from [6] is shown for comparisons of the other revised methods in fuzzy time series model, where our proposed method has the best performance.

3. **Forecasting Japanese Tourists for Taiwan Using the Proposed Method.** Over the past few decades, the Japanese tourists are Taiwan's major visitors in the tourism market because most Japanese are conscious of the closely historical and geographic relationship between Japan and Taiwan. Besides, the price level in Taiwan is still lower than Japan and the living environment is superior to some countries in the Asia, which attract many Japanese tourists to take a trip. The average annual tourists from Japan were above 500,000 over the period 1993-2008. Because the importance of the tourists from Japan, an accurate tourism demand forecasting can support much information to improve the tourism environment, including the linkage of scenic sites, developing new tourism spots, facility of transportation vehicles, development, training for the tourist local guides, and more supply of hotel graduates. For this purpose, we use the collected data from the Tourism Bureau of Republic of China, and forecast the annual Japanese tourists to Taiwan from 1993 to 2011 using our proposed model, where 1993-2006 is used for training data and 2007-2011 is used for testing data. The forecasting procedure is described as following steps.

Step 1. Define the universe of discourse for the historical data. In the first two column of Table 4, the minimum enrollment $D_{\min}$ is 323375 and the maximum enrollment $D_{\max}$ is 721351, then arbitrary chosen $D_1 = 375$ and $D_2 = 649$ derived $U = [323000, 722000]$.

Step 2. Define the fuzzy sets A_i on universe of discourse U. Because large numbers of tourists are from Japan in each year, we cut the intervals per 1000 in order to obtain more precise forecasting results [5]. Then the universe of discourse U is dividend into 399 intervals as $u_1 = [323000, 324000]$, $u_2 = [324000, 325000], \ldots, u_{399} = [721000, 722000]$.

Step 3. Establish the fuzzy logical relationships and group it. The fuzzy sets for tourist arrivals per year are shown in the 3-th column of Table 3 in which each fuzzy set has 399 elements. By analyzing the fuzzy logical relationships, we can obtain thirteen fuzzy logical relationship groups as shown in Table 2.

Step 4. Calculate the forecasted outputs. As in 2-th and 5-th columns of Table 2, group 1 to 12 are used for the training data, and their forecasting values are obtained in

Table 1. Comparisons of the forecasting enrollments with different revised methods

	Proposed method	**Song and Chissom [6]**	**Tsaur *et al.* [16]**	**Singh [17]**	**Cheng *et al.* [10]**	**Chen and Chung [15]**
MAPE (%)	0.49512	3.22376	1.87481	1.71204	2.08722	3.11005
RMSE	98.92944	650.405	367.3187	340.5475	438.2768	638.363

Table 2. The fuzzy logical relationship groups and the forecasting values

Fuzzy relationship groups		Forecasting value	Fuzzy relationship groups		Forecasting value
Group 1	$A_{161}, A_{252} \to A_{304}$	626500	Group 10	$A_{265}, A_1 \to A_{138}$	460500
Group 2	$A_{252}, A_{304} \to A_{283}$	605500	Group 11	$A_1, A_{138} \to A_{355}$	677500
Group 3	$A_{304}, A_{283} \to A_{253}$	575500	Group 12	$A_{138}, A_{355} \to A_{399}$	721500
Group 4	$A_{283}, A_{253} \to A_{175}$	497500	Group 13	$A_{355}, A_{399} \to \#$	718500
Group 5	$A_{253}, A_{175} \to A_{159}$	481500	Group 14	$A_{399}, A_{396} \to \#$	718500
Group 6	$A_{175}, A_{159} \to A_{205}$	527500	Group 15	$A_{396}, A_{396} \to \#$	718500
Group 7	$A_{159}, A_{205} \to A_{269}$	591500	Group 16	$A_{396}, A_{396} \to \#$	718500
Group 8	$A_{205}, A_{269} \to A_{265}$	587500	Group 17	$A_{396}, A_{396} \to \#$	718500
Group 9	$A_{269}, A_{265} \to A_1$	323500			

TABLE 3. Sensitivity analysis for value W

$\boldsymbol{W}$	θ	$\boldsymbol{F}$**(2007)**	$\boldsymbol{W}$	θ	$\boldsymbol{F}$**(2007)**	$\boldsymbol{W}$	θ	$\boldsymbol{F}$**(2007)**
1	0.5	699500	6	0.857	715214.29	11	0.916	717833.33
2	0.67	706833.33	7	0.875	716000	12	0.923	718115.39
3	0.75	710500	8	0.889	716611.11	13	0.929	718357.14
4	0.8	712700	9	0.9	717100	14	0.933	718566.67
5	0.833	714166.67	10	0.909	717500	15	0.9375	718750

TABLE 4. Adjusted tourist forecasting

Year	Actual tourist	Fuzzified tourist	Forecasting value	Residual	Adjusted residual	Adjusted forecasting value
1993	483481	A_{161}				
1994	574323	A_{252}				
1995	626152	A_{304}	626500	–348	–396	626104
1996	605673	A_{283}	605500	173	247	605747
1997	575613	A_{253}	575500	113	59	575559
1998	497928	A_{175}	497500	428	433	497933
1999	481544	A_{159}	481500	44	54	481554
2000	527074	A_{205}	527500	–426	–452	527048
2001	591081	A_{269}	591500	–419	–392	591108
2002	587170	A_{265}	587500	–330	–340	587160
2003	323375	A_1	323500	–125	–130	323370
2004	460231	A_{138}	460500	–269	–215	460285
2005	677937	A_{355}	677500	437	362	677862
2006	721351	A_{399}	721500	–149	–100	721400
2007(f)	737638		718500		93	718593
2008(f)	674506		718500		–274	721226
2009(f)	662644		718500		–442	721058
2010(f)	701561		718500		–321	721179
2011(f)	902733		718500		75	721575

TABLE 5. Comparisons among the forecasting methods

	ESM	Quadratic regression	Cubic regression	ARIMA (1,0,2)	AR(1)	Proposed model
1993∼2006 MAPE (%)	16.09	14.34	11.75	10.83	13.82	0.0062
2007∼2011 MAPE (%)	19.49	9.02	96. 80	27.04	21.51	8.24

3-th and 6-th columns of Table 2. However, we cannot obtain the extrapolative values from 2007 to 2011 by the fuzzy logical relationships because we have no information to judge what the next group is. For example, the forecasting tourists in 2007 is shown in Group 13 as well as the fuzzy logical relationship A_{355}, $A_{399 \rightarrow}\#$, which means that we use the tourists of 2005 and 2006 to forecast unknown tourists in 2007 by the relationship $F(2005)$, $F(2006) \rightarrow F(2007)$. We confirm that the tourists in 2007 will be increasing but deeply depend on 2005 and 2006, and thus a larger W should be chosen. We choose θ to be larger than or equal to 0.9, without loss its generality, as testing each value in Table 3, we choose $W = 12$ to forecast the Japanese tourists in 2007. Clearly, forecasting value

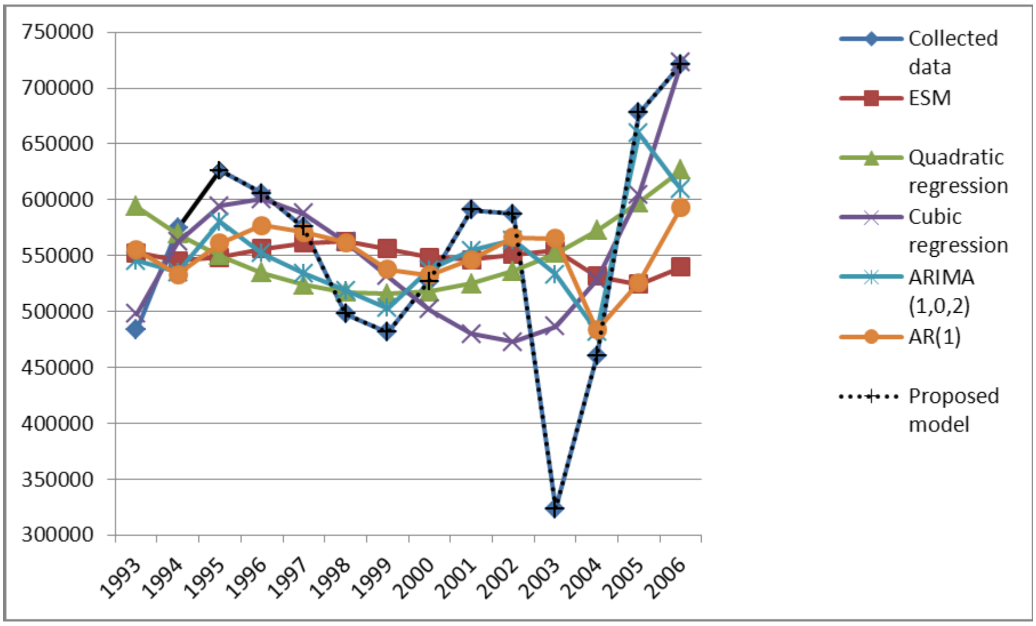

FIGURE 1. Comparisons among the forecasting methods

of 2007 is fitted to A_{396} whose defuzzy value is 718500 as in the Group 13. Next, for the forecasting value of 2008, we use the fuzzy set A_{396} derived forecasting tourists in 2007; then the fuzzy logical relationship for the forecasting value of 2008 is shown in Group 14 as well as A_{399}, $A_{396 \rightarrow} \#$. By the same trend, we can forecast the Japanese tourists in 2008 as 718731. According to the above result, we can fit the forecasting value of 2008 to be a fuzzy set A_{396} and obtain its defuzzy value as 718500 shown in Table 2 of Group 14. Consequently, in Table 2, the forecasting value of 2009, 2010, and 2011 can be obtained as Groups 15, 16 and 17, respectively.

Step 5. Fine-tune the forecasting values. We take residuals from 1995 to 2006 in the 5-th column of Table 4 into Equation (6), the parameters a_0, a_i and b_i in Fourier series are solved, $i = 1, 2, \ldots, 5$. Then, we can obtain the adjusted residual shown in 6-th column of Table 4, where the adjusted forecasting values are obtained in 7-th column.

Next, we use the derived Fourier series to forecast the adjusted residual from 2007 to 2011 in the 6-th column of Table 4, and then their adjusted forecasting values can be easily obtained in 7-th column. In Table 5 and Figure 1, performance comparisons among the forecasting models are illustrated, including exponential smoothing model (ESM), quadratic regression, cubic regression, auto-regression AR(1), ARIMA(1, 0, 2), and our proposed model form 1993 to 2006 in which our proposed model has the best performance MAPE = 0.0062% in the in-sample prediction. Moreover, in Table 5 for the out-of-sample extrapolation, our proposed model owns the best performance from 2007 to 2011.

4. **Conclusions.** In this study, we first propose a fuzzy time series model using Fourier series for adjusting the residual terms, and then we obtain the adjusted forecasting values with smaller forecasting errors. Second, we illustrate the proposed method to the application for Japanese tourists to Taiwan, where the forecasting results indicated considerable forecasting performance. Both the forecasting results and the analysis confirm the potential benefits of the new approach in terms of tourism demand forecasting. Most importantly, the forecasting performance in the illustration is archived with smaller MAPE, and shows a better forecasting error on the out-of-sample. As a result, the proposed model forecasts with smaller forecasting error than some statistical models with little

data. If the fuzzy time series model with residual analysis using Fourier series method meets its expectations, then this approach will be easily applied to be an important tool in forecasting limited data.

Acknowledgment. The authors gratefully acknowledge the financial support from National Science Foundation with project No. NSC 97-2221-E-032-050.

REFERENCES

[1] H. Qu and S. Lam, A travel demand model for Mainland Chinese tourists to Hong Kong, *Tourism Management*, vol.18, no.8, pp.593-597, 1997.

[2] C. Lim and M. McAleer, Time series forecasts of international travel demand for Australia, *Tourism Management*, vol.23, no.4, pp.389-396, 2002.

[3] H. Song and S. F. Witt, Forecasting international tourist flows to Macau, *Tourism Management*, vol.27, no.2, pp.214-224, 2006.

[4] C. Lim, M. McAleer and J. C. H. Min, ARMAX modelling of international tourism demand, *Mathematics and Computers in Simulation*, vol.79, no.9, pp.2879-2888, 2009.

[5] K. M. Cheng, Tourism demand in Hong Kong: Income, prices, and visa restrictions, *Current Issues in Tourism*, vol.15, no.3, pp.167-181, 2012.

[6] Q. Song and B. S. Chissom, Forecasting enrollments with fuzzy time series – Part I, *Fuzzy Sets and Systems*, vol.54, no.1, pp.1-9, 1993.

[7] Q. Song and B. S. Chissom, Forecasting enrollments with fuzzy time series – Part II, *Fuzzy Sets and Systems*, vol.62, no.1, pp.1-8, 1994.

[8] L. W. Lee, L. H. Wang and S. M. Chen, Temperature prediction and TAIFEX forecasting based on fuzzy logical relationships and genetic algorithms, *Expert Systems with Applications*, vol.33, no.3, pp.539-550, 2007.

[9] I. H. Kuo, S. J. Horng, T. W. Kao, T. L. Lin, C. L. Lee and Y. Pan, An improved method for forecasting enrollments based on fuzzy time series and particle swarm optimization, *Expert Systems with Applications*, vol.36, no.3, pp.6108-6117, 2009.

[10] C. H. Cheng, T. L. Chen, H. J. Teoh and C. H. Chiang, Fuzzy time-series based on adaptive expectation model for TAIEX forecasting, *Expert System with Applications*, vol.34, no.2, pp.1126-1132, 2008.

[11] H. L. Wong, Y. H. Tu and C. C. Wang, Application of fuzzy time series models for forecasting the amount of Taiwan export, *Expert Systems with Applications*, vol.37, no.2, pp.1465-1470, 2010.

[12] C. H. Wang, Predicting tourism demand using fuzzy time series and hybrid grey theory, *Tourism Management*, vol.25, no.3, pp.367-374, 2004.

[13] C. H. Wang and L. C. Hsu, Constructing and applying an improved fuzzy time series model: Taking the tourism industry for example, *Expert Systems with Applications*, vol.34, no.4, pp.2732-2738, 2008.

[14] Y. H. Lin and P. C. Lee, Novel high-precision grey forecasting model, *Automation in Construction*, vol.16, no.6, pp.771-777, 2007.

[15] S. M. Chen and N. Y. Chung, Forecasting enrollments using high-order fuzzy time series and genetic algorithms, *International Journal of Intelligent Systems*, vol.21, no.5, pp.485-501, 2006.

[16] R. C. Tsaur, J. C. O. Yang and H. F. Wang, Fuzzy relation analysis in fuzzy time series model, *Computer and Mathematics with Applications*, vol.49, no.4, pp.539-548, 2005.

[17] S. R. Singh, A simple method of forecasting based on fuzzy time series, *Applied Mathematics and Computation*, vol.186, no.1, pp.330-339, 2007.