Forecasting International Tourism Demand in China Mainland Based on Comparison of Three Models

Mei Li
Economics School, Guangxi University, Nanning 530004, China.

Abstract: Tourism is an important part of the national economy, accuracy forecasting tourism demand is conducive to promoting the sustainable development of the tourism industry. In order to forecast international tourism demand in Mainland China, this paper uses the monthly tourist arrivals from Mainland China to Thailand, Japan and Korea time span from January 2011 to December 2019, and consider Baidu search engine as exogenous variable. Using three commonly used forecasting models, namely, seasonal autoregressive integrated moving average with exogenous variable (SARMAX) model, back propagation neural network (BPNN) model and support vector regression (SVR) model to long term and short term forecast the international tourism demand in Mainland China. The results show that the SARIMAX model generate highest prediction accuracy for almost all evaluation indicators and forecasting steps, while BPNN model and SVR model show different forecasting accuracy under different conditions, which provides guidance for the selection of forecasting models for tourism demand.
Keywords: SARIMAX Model; BPNN Model; SVR Model; Tourism Demand Forecasting

Introduction

Recently, driven by the continuous development of the economic level, the demand for leisure and entertainment is growing, and tourism is one of the most important components. The radiation scope of tourism includes catering, transportation, accommodation, entertainment and other consumer services, tourism plays an important role in economic growth, meanwhile, destination tourism management also faces many challenges, so accurate tourism demand prediction is particularly important (Li et al., 2016). Online search behavior reflects the tourists' attention, interest and potential consumption intention to some extent, and online search data provides a new data source for scientific research. The amount of online search queries is closely related to tourism demand, and is also an important variable in tourism demand forecast (Dinis et al., 2019). This study takes Baidu search engine as the main factor affecting tourism demand, compares the forecasting ability of SARIMAX model, BPNN model, and SVR model, so as to provide reference for the selection of tourism demand prediction models in the future. so as to offer guidance for destination tourism management.The contribution of this paper is mainly to compare the prediction performance of econometric models and artificial intelligence models that are widely used at present, so as to provide reference for the selection of prediction models and reduce the time cost brought by the selection of the optimal model.

1. Literature review

After decades of development, forecasting models have been gradually proposed and applied by scholars, and many achievements have been made in the field of tourism demand forecasting, the current forecasting models are mainly divided into four categories: time series model, econometric model, artificial intelligence algorithm and hybrid model (Song et al., 2019).

The time series models such as ETS model, Seasonal Naïve model and seasonal autoregressive integrated moving average

(SARIMA) model generate forecasting results based on historical data, which reduces the cost of data collection because only one variable is needed to build the model, and it is widely used for benchmark models in tourism demand forecasting. When exogenous variables are introduced into the forecasting model, the econometric model is used, when forecasting tourism demand, some factors affecting tourism demand are taken into account in the prediction model, such as own price, substitute prices and consumer's income (Song et al., 2003; Wong et al., 2007), Tsui & Balli (2017) taken economic conditions, flight services, fluctuations and shocks as exogenous variable to forecast international passenger arrivals for the eight key Australian airports, they found that SARIMAX model performs better than SARIMA model. Prilistya et al.(2021) add COVID-19 and Google trend data to SARIMAX model to forecast international tourist arrivals in Indonesia, result display that the best prediction results are generated by SARIMAX model. Nontapa et al.(2020) proposed decomposition method with SARIMAX models, after analyzing international tourists in Thailand, they found the proposed method has the lowest average MAPE for 3 months anr 12 months. Lee & Choi (2020) introduce the 9/11 attack, the SARS and MERS epidemic, and the deployment of THAAD as explanatory variables, and applied SARIMAX model to forecast the number of inbound foreigners visiting Korea, this study confirmed the prediction performance of the SARIMAX model. Ampountolas (2021) put temperature, holidays, competitive set ranking in the SARIMAX model for forecasting daily hotel demand, the result show that SARIMAX model outperformed the Artificial Neural Network-Multilayer Perceptron (ANN-MLP) model, SNaïve, ARIMA and GARCH model in every one horizon except one out of seven forecast horizons. Park et al. (2021) used news media to forecast monthly tourist arrivals from mainland China and the US, the findings revealed that SARIMAX model has lower prediction error than other benchmark models. Among all AI models, BPNN is the most used (Wu et al., 2017), the BPNN model and SVR model is also proved to have a good prediction effect (Li et al., 2018; Cho, 2003; Akin, 2015). Additionally, other AI models were also used, such as Multilayer Perceptron (Claveria et al., 2015),Grey model (Sun et al., 2016) and Random Forest (RF) model (Feng et al., 2019). With the deepening of research about tousism demand forecasting, many scholars have proposed hybrid models, which combines two or more statistical and/or artificial intelligence (AI) techniques to obtain a result (Fajardo-Toro et al., 2019), there mix different models, such as hybridisation of the SARIMA model with the SVR model (Abellana et al., 2021; Cang, 2014); the ARIMA model with a Back Propagation Neural Network (BPNN) model (Liao et al., 2013), they point out that hybrid model shows better performance than individual model, while some researchers found that hybrid models are are not guaranteed to boost prediction accuracy (Oh & Morzuch, 2005; Shen et al., 2010).

With the application and development of Internet technology, tourists are used to searching for tourism related keywords on online platforms such as Baidu Index before traveling. Therefore, Baidu Search Index contains important information about tourism demand, which can be included in the study of tourism demand, Bangwayo-Skeete & Skeete (2015) adopted MIDAS model to verify the role of online search engine in tourism demand forecasting, after that, many researches introduced search engine data into prediction and it has been proved to have an effect on tourism demand (Liu et al., 2021; Wen et al., 2021).

Many achievements have been made in tourism demand models, and many types of variables have also been considered. However, it is particularly important to compare the prediction effects based on different models, which will provide strong guidance and reference for selecting prediction models, to achieve this, this study compares SARIMAX model, BPNN model and SVR model for forecasting tourist arrivals from Mainland China to Thailand, Japan and Korea.

2. Methodology

In order to empirically analyze the forecasting effect of three models, this study build a research framework, which are mainly divided into four steps. The first step is data collection, which mainly includes the tourists arrivals from mainland China to Thailand, Japan and Korea and Baidu search engine data. The second step is to select Baidu search keywords using correlation analysis. The third step is to build forecasting model including to SARIMAX, BPNN and SVR model. The last step is to evaluate the forecasting accuracy, the evaluation index adopted is mainly Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE). Three prediction models used in this study will be introduced as follows.

2.1 SARIMAX model

The SARIMAX model is widely used (Ampountolas, 2021; Prilistya et al., 2021), the SARIMAX model can be expressed as:

$$\Phi(B^m)\varphi(B)(1-B^m)^D(1-B)^d y_t = c + \sum_i \sum_k \beta_{i(t-k)} X_{i(t-k)} + \Theta(B^m)\theta(B)\varepsilon_t$$

where B is the backshift operator; m denotes the seasonal cycle; $\Phi(x)$ and $\Theta(x)$ are polynomials of orders P and Q, respectively; $\varphi(x)$ and $\theta(x)$ are polynomials of orders p and q, respectively; ε_t is a white noise process with a mean of zero and a variance of σ^2; and d and D refer to the rank of difference, which is determined by a unit root test, X_i represents the ith exogenous variable, which is composite index of Baidu index in this study and the corresponding coefficient.

2.2 BPNN model

Back propagation is a learning mode that requires supervised learning, which is mainly reflected in the training process of BPNN. BPNN usually consists of an input layer, one or more hidden layers and an output layer. The structure of BPNN network with only one hidden layer adopted in this study. The input layer is mainly used to obtain input information, and the number of neurons depends on the number of input variables; The hidden layer is mainly used for feature extraction and weight adjustment to achieve the optimization of training; The output layer is used to interface with the hidden layer and output the results, and adjust the weights to form correct responses to different hidden layer neuron stimuli.

The BPNN model consists of forward propagation and backward propagation. Its basic idea is to calculate the weight value of each variable and each layer through the forward propagation of the signal, and then adjust the variables and the weight of each layer through the back propagation to make the model optimal. BPNN model has been proved to be able to capture the nonlinear relationship between variables, it was used in power forecasting, stock forecasting, air pollution index forecasting, etc, but different studies have drawn different conclusions.

2.3 SVR model

Support vector machine a machine learning method that has been used for prediction (Lin et al., 2013). At first, it was only used for classification problems, its basic idea is to make all sample points approach the hyperplane in the process of finding the optimal hyperplane, so that the total deviation of sample points from the hyperplane can be minimized. Support Vector Regression (SVR) is the application and expansion of support vector machine in regression estimation. Its basic idea is to transplant the hinge loss function of support vector machine to the regression problem. Record the regression function (hyperplane) is $f(X) = \beta_0 + X^{'}\beta$, and use this function to predict the continuous response variable y. The objective function of SVR is mathematically expressed as follows:

$$\min_{\beta,\beta_0} \frac{1}{2}\beta^{'} + C\sum_{i=1}^{n} \ell_{\varepsilon}(y_i - f(X_i))$$

where, C>0 is the regularization parameter, $\mathrm{Z} = y_i - f(X_i)$ is the residual, $\ell_{\varepsilon}(.)$ is the insensitive loss function, and is defined as follows:

$$\ell_{\varepsilon}(\mathrm{Z_i}) = \begin{Bmatrix} 0, & \text{if } |\mathrm{Z_i}| \le \epsilon \\ |\mathrm{Z_i}| - \varepsilon, & \text{if } |\mathrm{Z_i}| > \epsilon \end{Bmatrix}$$

where, $\varepsilon > 0$ is also an adjustment parameter. For the solution of SVR objective function, relaxation variables can be introduced and solved by Lagrangian function. The relaxation variable is introduced below ζ_i , ζ_i^*,

$$\min \frac{1}{2}\beta^{'}\beta + \mathrm{C}\sum_{i=1}^{\ell}(\zeta_i + \zeta_i^*)$$

$$\text{s.t.} \quad y_i - \beta^{'}X_i - b \le \varepsilon + \zeta_i$$

$$\beta^{'}X_i + b - y_i \le \varepsilon + \zeta_i^*$$

$$\zeta_i\zeta_i^* \ge 0$$

By introducing Lagrangian multipliers and using Lagrangian functions and duality principle, can obtain:

$$\max Z = -\frac{1}{2}\sum_{i,j=1}^{\ell}(a_i^* - a_i)(a_j^* - a_j)(X_i, X_j) - \varepsilon\sum_{i=1}^{\ell}(a_i^* + a_i) + \sum_{i=1}^{\ell}y_i(a_i^* - a_i)$$

$$\sum_{i=1}^{n}(a_i^* - a_i) = 0$$

$$\text{s.t.} \quad 0 \le a_i \le C$$

$$0 \le a_i^* \le C$$

where a_i is a Lagrange multiplier and non negative, according to the above, can obtain: $\beta = \sum_{i=1}^{n}(a_i - a_i^*)X_i$

use KKT (Karush Kuhn Tucker) condition to calculate the deviation b, and finally get the expression of function $f(X)$: $f(X) = \sum_{i=1}^{\ell}(a_i - a_i^*)K(X, X_i) + b$, where, $K(X, X_i) = \Phi'(X)\Phi(X_i)$ is a kernel function satisfying Mercer condition.

3. Data

After collecting the data released by official tourism bureaus, authoritative media, and tourism associations, according to the number of outbound tourists, Thailand, Japan and Korea are ranked in order, so these three countries are selected as the main analysis objects of this study. This study collect monthly tourist arrivals from mainland China to Thailand, Japan and Korea from January 2011 to December 2019, correspondingly, this study obtain daily Baidu search engine data from January 1, 2011 to December 31, 2019, and the daily Baidu search engine data is processed into monthly data by moving average (Bokelmann & Lessmann,2019). The correlation between monthly tourists arrivals and Baidu search keywords is shown in Table 1, the bold keywords are the 10 most relevant keywords. From the results of correlation analysis, Baidu search keywords with high relevance to tourists arrivals are mainly reflected in tour, while the relationship between dining and shopping and tourism demand is relatively low, which reflects the concerns of tourist arrivals entering Korea to some extent.

Table 1. Correlation between search query keywords and tourists arrivals

Search query keywords	Lag order	Correlation	Search query keywords	Lag order	Correlation
Dining			Is Korea fun	1	0.554
Korean course	0	0.400	Korea travel	1	0.401
Korean Fried Rice Cake	8	0.620	Korea tourism	1	0.274
Korean BBQ	9	0.656	Korea travel quotation	7	-0.153
Korean cuisine	12	0.398	Korea tourist map	1	0.650
Korean dishes	2	0.592	How much is the Korean tour	0	0.680
Korean Kimchi	8	0.645	Korea tourism development bureau	12	0.229
Korean food	12	0.478	Korea travel price	0	-0.159
Korean Recipe	0	-0.360	Korea tourism introduction	0	-0.197
Korean Sushi	4	0.229	Korean tourist group	1	0.389
Korean specialty	0	0.413	Notes on Korea tourism	1	0.068
Korean Fried Noodles	2	0.674	Seoul tourism, Korea	0	0.160
Traffic			Korea weather forecast	0	0.215
Seoul Airport	0	0.829	Korean journey	12	0.217
Map of Korea	0	0.857	Korea time	0	0.530
Korean visa	0	0.776	Korea free travel	12	0.491
Korea Metro	3	0.395	Korea free travel strategy	0	0.620
Chinese version of Korean map	0	0.412	Jeju island travel	1	0.479
Korean tourist visa	6	0.580	Travel to Korea	3	-0.074

Incheon International Airport	12	0.498	**Attraction**		
Busan Airlines	12	0.575	**Korea seoul Tower**	0	0.802
Korean aircraft	0	0.518	**Korea busan**	0	0.755
Korean car rental	0	0.665	**Korea Jeju island**	1	0.742
Lodging			Takashimaya	0	0.511
Korean hotel	0	0.680	Korea Ancheng	12	0.563
Seoul hotel	0	0.450	Korea daegu	0	0.682
Shopping			East Gate of South Korea	0	0.505
Korea shopping	1	0.403	Korea fanyu temple	6	-0.095
Korea shopping Strategy	1	0.522	Korea Busan tourist attractions	6	-0.105
Korean Price	1	0.517	Korea sea yuntai	0	0.276
Korean won exchange rate to RMB	0	0.193	South Korea	10	0.342
Tour			Korea Jingfu Palace	0	0.630
Korean weather	0	0.834	Korean tourist attractions	0	0.407
Seoul tourism	1	0.748	Korea Mingdong	0	0.586
Busan tourism	0	0.715	Korea Incheon	0	0.471
Seoul tourism	1	0.748	Hanna Mountain	12	0.267
Busan tourism	0	0.715	Lotte world	0	0.534
Korea tourism strategy	0	0.694			

4. Empirical results

In order to compare SARIMAX model, BPNN model and SVR model,, RMSE, MAE and MAPE are used as the evaluation index of forecasting performance, the formula of the three evaluation indicators is as follows:

$$RMSE=\sqrt{\frac{1}{N}\sum_{i=1}^{N}|y_i-y_i'|^2} \qquad MAE=\frac{1}{N}\sum_{i=1}^{N}|y_i-y_i'| \qquad MAPE=\frac{1}{N}\sum_{i=1}^{N}|\frac{y_i-y_i'}{y_i}|$$

where y_i denotes the actual value of the observation; y_i' is the prediction value; and N represents the length of the forecasting period, M indicates the number of seasons (for example,when the data frequency is monthly, M =12).

Data from January 2011 to December 2018 is used for training set, and the remaining data is used for the test set in the process of constructing the SARIMAX model, BPNN model and SVR model, and the extending method is used in this study. Since the lag period is considered in the calculation of correlation, the final data is from February 2011 to December 2019 for Korea and Japan. Take one step ahead as an example, the tourists arrivals data from March 2011 to December 2018 and the Baidu search engine February 2011 to November 2018 are used as the initial training sets to forecast the tourist arrivals demand of January 2019; the window is then extended by one month each time until all 12 month forecasts (January 2019 to December 2019) are generated exhaustingly. Through the above modelling process, one-, two-, three-, six-, nine- and twelve-step-ahead forecasts are generated by the SARIMAX model, BPNN model and SVR model, respectively, and RMSE, MAE, MAPE are taken as forecasting accuracy measures. When building the BPNN model, in addition to the training set and test set, this study also sets the validation set for filtering out the best parameters, and the minimum RMSEs value is taken as the screening standard.

Table 2 lists the forecasting effect evaluation index of the three models, the results reveal that the SARIMAX model yields smaller prediction error than BPNN model and SVR model for almost all evaluation indicators and forecasting steps, meanwhile, the BPNN model and SVR model are different forecasting effects under different destinations and forecasting step, specially, for Korea, SVR model performs better than BPNN model in all forecasting steps, while for Thailand, this conclusion is only true in short-term forecasts, for Japan, BPNN model is outperform SVR model in general, generally, SARIMAX model has ideal prediction effect. The results imply that each model has its own advantages and characteristics, and can show perfect forecasting ability under specific circumstances, which are consistent with previous studies (Oh & Morzuch, 2005; Shen et al., 2011).

Table 2. Forecasting accuracy valuation

	Kerea			Thailand			Japan		
Horizon	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE
<u>*h = 1*</u>									
SARIMAX	2733	2428	8.39	2658	2453	14.80	1769	1398	5.46
BPNN	2980	2345	14.04	3067	2511	8.79	4576	3613	13.93
SVR	1944	1783	10.84	3448	3072	10.15	5405	4206	15.30
<u>*h = 2*</u>									
SARIMAX	2276	2010	6.90	1165	921	5.67	1975	1687	6.38
BPNN	2766	2430	15.38	4544	3419	10.60	2685	1964	7.28
SVR	1618	1283	7.92	4654	3350	10.33	5692	4024	14.10
<u>*h = 3*</u>									
SARIMAX	1920	1584	5.11	2431	2172	13.44	2239	1790	6.99
BPNN	2647	1916	12.19	4398	3365	10.58	3893	3274	12.47
SVR	1518	1257	7.52	4909	3616	11.08	5825	4780	17.45
<u>*h = 6*</u>									
SARIMAX	3567	2689	8.96	2613	1833	11.01	2135	1824	7.24
BPNN	2248	1881	11.19	7288	5562	17.71	3458	2514	8.85
SVR	1879	1487	9.00	4760	3424	10.55	4421	3408	12.14
<u>*h = 9*</u>									
SARIMAX	3736	2376	7.86	1587	1216	7.54	2333	1678	6.64
BPNN	1901	1664	10.00	5241	4796	16.24	3914	2852	9.92
SVR	2015	1682	10.21	3693	2589	8.26	5159	3836	13.37
<u>*h = 12*</u>									
SARIMAX	2729	2243	7.40	1466	1251	7.79	2016	1625	6.30
BPNN	1649	1308	7.60	4295	3398	11.11	4634	3556	13.03
SVR	1318	1104	6.67	5360	4160	12.82	3931	2856	9.98

5. Conclusion

Under the background of computer technology development and big data era, internet search behavior has gradually spawned many data sources, which provides available variables for academic research. Currently, a large body of literature suggests that Baidu search engine index can generate positive effect on the boost prediction accuracy of tourism demand. Various prediction models have become the focus of scholars, and have been proved that different prediction models have their own advantages and characteristics. This study takes the number of tourists arrivals from Mainland China to Korea, Thailand and Japan as the empirical research object, and selects Baidu search engine as the main factor affecting tourism demand, first, this study obtain data including Baidu search keywords and tourist arrivals, then correlation analysis is used to eliminate some keywords without predictive power, the ten most relevant keywords are used as the final keywords, which are put into three models as exogenous variables. Through empirical test, the results demonstrate that among the three models, SARIMAX model shows the optimal forecasting ability, while BPNN model and SVR model show different prediction performance in different situations, this imply that SARIMAX model can be widely used for forecasting.

The main defects of this study are: first, other factors that affect tourism demand, such as macroeconomic variables such as price and income, weather, emergencies and other dummy variables are not included; Second, some other econometric model, artificial intelligence algorithms and deep learning methods, such as multi-layer perceptron and short-term memory neural network are not used in this study, it is necessary to mention that Baidu search index and target variables always have different frequencies, and the use of MIDAS model is particularly effective, which can be involved in future research.

References

[1] Abellana, D. P. M., Rivero, D. M. C., Aparente, M. E., & Rivero, A. (2021). Hybrid SVR-SARIMA model for tourism forecasting using PROMETHEE II as a selection methodology: A Philippine scenario. *Journal of Tourism Futures. 7*(1), 78-97.

[2] Akin, M. (2015). A novel approach to model selection in tourism demand modeling. *Tourism Management, 48*, 64-72.

[3] Ampountolas, A. (2021). Modeling and Forecasting Daily Hotel Demand: A Comparison Based on SARIMAX, Neural Networks, and GARCH Models. *Forecasting, 3*(3), 580-595.

[4] Bangwayo-Skeete, P.F., & Skeete, R.W. (2015). Can Google data improve the forecasting performance of tourist arrivals? Mixed-data sampling approach. *Tourism Management, 46*, 454-464.

[5] Bokelmann, B., & Lessmann, S. (2019). Spurious patterns in Google Trends data-An analysis of the effects on tourism demand forecasting in Germany. *Tourism management*, *75*, 1-12.

[6] Cang, S. (2014). A comparative analysis of three types of tourism demand forecasting models: Individual, linear combination and non-linear combination. *International Journal of Tourism Research, 16*(6), 596-607.

[7] Cho, V. (2003). A comparison of three different approaches to tourist arrival forecasting. *Tourism management*, *24*(3), 323-330.

[8] Claveria, O., Monte, E., & Torra, S. (2015). Tourism Demand Forecasting with Neural Network Models: Different Ways of Treating Information. *International Journal of Tourism Research, 17*(5), 492-500.

[9] Dinis, G., Breda, Z., Costa, C., & Pacheco, O. (2019). Google Trends in tourism and hospitality research: a systematic literature review. *Journal of Hospitality and Tourism Technology*.

[10] Fajardo-Toro, CH., Mula, J., & Poler, R. (2019). Adaptive and hybrid forecasting models—A review. *Engineering Digital Transformation,* 315–322.

[11] Feng, Y., Li, G., Sun, X., & Li, J. (2019). Forecasting the number of inbound tourists with Google Trends. *Procedia Computer Science*, *162*, 628-633.

[12] Lee, G. C., & Choi, S. H. (2020). Forecasting Foreign Visitors using SARIMAX Models with the Exogenous Variable of Demand Decrease. *Journal of the Society of Korea Industrial and Systems Engineering, 43*(4), 59-66.

[13] Li, S., Chen, T., Wang, L., & Ming, C. (2018). Effective tourist volume forecasting supported by PCA and improved BPNN using Baidu index. *Tourism Management*, *68*, 116-126.

[14] Li, X., Wu, Q., Peng, G., & Lv, B. (2016). Tourism forecasting by search engine data with noise-processing. *African Journal of Business Management*, *10*(6), 114-130.

[15] Liao, Z., Jin, M., Luo, Y., Ren, P., & Gao, H. (2013). Research on prediction of tourists' quantity in Jiuzhaigou Valley scenic based on ABR@ G integration model. *International Journal of Environment and Pollution, 51*(34), 176–191.

[16] Lin, K. P., Pai, P. F., Lu, Y. M., & Chang, P. T. (2013). Revenue forecasting using a least-squares support vector regression model in a fuzzy environment. *Information Sciences, 220*, 196–209.

[17] Liu, H., Liu, Y., Li, G., & Wen, L. (2021). Tourism demand nowcasting using a LASSO-MIDAS model. *International Journal of Contemporary Hospitality Management*.

[18] Nontapa, C., Kesamoon, C., Kaewhawong, N., & Intrapaiboon, P. (2020, November). A new time series forecasting using decomposition method with SARIMAX model. *In International Conference on Neural Information Processing* (pp. 743-751). Springer, Cham.

[19] Oh, CO., & Morzuch, BJ. (2005). Evaluating Time-Series Models to Forecast the Demand for Tourism in Singapore: Comparing Within-Sample And Postsample Results. *Journal of Travel Research*, *43*(4), 404–413.

[20] Park, E., Park, J., & Hu, M. (2021). Tourism demand forecasting with online news data mining. *Annals of Tourism Research, 90*, 103273.

[21] Prilistya, S. K., Permanasari, A. E., & Fauziati, S. (2021, August). The Effect of The COVID-19 Pandemic and Google Trends on the Forecasting of International Tourist Arrivals in Indonesia. In 2021 IEEE Region 10 Symposium (TENSYMP) (pp. 1-8). IEEE.

[22] Shen, S., Li, G., & Song, H. (2011). Combination forecasts of international tourism demand. *Annals of Tourism Research*, *38*(1), 72-89.

[23] Song, H., Qiu, R. T., & Park, J. (2019). A review of research on tourism demand forecasting: Launching the Annals of Tourism Research Curated Collection on tourism demand forecasting. *Annals of Tourism Research*, 75, 338-362.

[24] Song, H., Wong, K. F., & Chon, K. S. (2003). Modelling and forecasting the demand for Hong Kong tourism. *International Journal of Hospitality Management, 22*, 435–451.

[25] Sun, X., Sun, W., Wang, J., Zhang, Y., & 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.

[26] Tsui, W. H. K., & Balli, F. (2017). International arrivals forecasting for Australian airports and the impact of tourism marketing expenditure. *Tourism Economics, 23*(2), 403-428.

[27] Wen, L., Liu, C., Song, H., & Liu, H. (2021). Forecasting tourism demand with an improved mixed data sampling model. *Journal of Travel Research*, *60*(2), 336-353.

[28] Wong, K. K., Song, H., Witt, S. F., & Wu, D. C. (2007). Tourism forecasting: to combine or not to combine?. *Tourism management, 28*(4), 1068-1078.

[29] Wu, D., Song, H., & Shen, S. (2017). New developments in tourism and hotel demand modeling and forecasting. *International Journal of Contemporary Hospitality Management, 29*(1), 507-529.