JOURNAL OF H&T
www.johat.org

The Predictive Power of Google searches to Forecast Inbound Tourism in China

Huiyue Liu, Gege Zhang
Shenzhen Tourism College,
Jinan University, Shenzhen,Guangdong,,China

Abstract: Development of a scientific method to predict inbound tourism demand is necessary for the effective allocation of resources and ability to rationally formulate a national tourism strategy. This study provides evidence that the use of internet search data can help China to accurately predict the amount of inbound tourism demand. Here, we compared the prediction ability ofan autoregressive moving-average (ARMA) model and anautoregressive distribution lag (ARDL) model. We found that the ARDL model offers a better fit andhas higher prediction accuracy than the ARMA model, andsearch engine data can predict inbound tourism demand accurately. The studydemonstrates the value and predictive power of search engine data and provides a new framework for using internet search query data. These analyses also dynamically evaluate the information demand of inbound tourists and can be used to guide policy makers on how to useinternet search datawhen making decisions.

Keywords: Google Trends; Inbound Tourists; ARMA Models; ARDL Model

Introduction

Inbound tourism plays an important role in China's tourism industry. Inbound tourism positively contributes to foreign exchange earnings, provides employment, and stimulates economic development. Tourismdemand is also a reflection of tourism competitiveness, cultural soft power, and even the degree of opennessof a country or region(Ma & Li, 1999). Many scholarshave studied the development of inbound tourism in China from different perspectives. Among them, thescientific and reasonable prediction of inbound tourism demand is directly related to the formulation andimplementation of China's tourism development strategy(Song & Li, 2008). Traditional prediction methods use statistical data obtained from governmental and relevant statisticaldepartments. However, these data are collected over a long period of time with low frequency. Furthermore,they often lack of timeliness in the release of statistical results, and the data size have great limitations foruse in a prediction model. The use of high quality data resources is essential for prediction accuracy(Huanget al., 2013). As information technology advances, internet search queries have become a way for travelers tofind information about their destinations. Inbound tourists use search engines to research tourist attractions,politics, economy, culture, and other tourism-related information in order to plan their trips. Search enginesgenerate a large amount of user data and provide valuable tourist information in terms of their interests,needs and feedback. For this reason, search engines are considered a prime source of information regardingtourism demand in the era of big data (Li et al., 2017). In addition, the immediacy of internet search datamakes up for the lag of traditional prediction methods. Baidu and Google have launched the Baidu Indexand Google Trends analysis capabilities, respectively, which we can use to observe trends in topics andmeasure search volume for a keyword during a specific period of time. These metrics directly reflect socialhotspots as well as user interests and demands(Huang et al., 2013). Currently, Chinese scholars mainlyuse the Baidu index data to predict tourist numbers. The Baidu index can effectively measure domestictourists' searches for tourism information. However, the Google search engine captures more than 90% ofthe global market and can therefore more fully reflect potential inbound tourist queries[1]. For this reason, itmay represent a more scientific mechanism for predicting the number of inbound tourists. Therefore, in thispaper, we extracted keywords from Google search trend data to evaluate foreign tourists queries for tourisminformation about China. Using this data, we constructed an autoregressive moving-average (ARMA) modeland an autoregressive distribution lag (ARDL) model. We used these two models to simulate the volume ofinbound tourists.

In this paper, we aim to identify the focus of internet searches from a scientific perspective. We willthen extract the time series data to determine whether internet search query volume can effectively predictinbound tourism demand. We will also build a dynamic model to identify the type of information touristsseek when traveling to China. This paper proceeds as follows: part 2 reviews the relevant literature; part 3describes the design of the prediction model; part 4 introduces the empirical analysis and the research results;part 5 conducts the in sample and out-of-sample predictions; part 6 summarizes our main conclusions.

Literature review

Factors that influence international tourism demand

As the study of inbound tourism grows increasingly in-depth, the factors which influence internationaltourism demand are gradually receiving greater attention in academic circles. The study of these factors canbe traced as far back as 1961 when Guthrie studied the impact of geographical location, trade advantages,and immigration on a country's tourism(Guthrie, 1961). The study confirmed that the country's affluenceand geographical location have the greatest impact on international tourism demand. Subsequent studieshave subdivided the factors that affect international tourism demand. One method is to classify the factorsusing push-pull theory. Push-pull theory was originally used to study the phenomenon of population andimmigration. Dann (1977) first applied push-pull theory to tourism research in order to explain the flow oftourists. Push factors are influencing factors that encourage people to travel. Several scholars have identifiedcommon push factors for specific tourist destinations, such as "escaping from everyday environment", "socialinteraction", "novelty", and

[1]A United States website traffic monitoring agency providing various types of statistical reports and website traffic statistics services:http://gs.statcounter.com/search-engine-market-share/all/china/.

"reputation"(Kozak, 2002; Botha et al., 1999; Oh et al., 1995). Pull factors refer to the factors which attract people to travel to a particular tourism destination. Current common pullfactors include tourist attractions, natural climatic conditions, physical environment, special folklore, localfoods and festivals (Paul C. Fakeye & John L. Crompton, 1991; Jeong & Park, 1997; Kim & Richardson,2003; Zheng et al., 2014). In addition to push and pull factors, the theory suggests that there are resistancefactors which can include supply capacity, travel time, and natural and man-made disasters (Frechtling,1996).

In addition to the factors identified by the push-pull theory, several studies have examined economic factors and non-economic factors that influence inbound tourism demand (Crouch, 1995; Lim, 1999). Amongthem, Crouch (1995) used meta-analysis to examine 80 influencing factors in order to identify the determinants for inbound tourism demand in destination countries. He found that the main factors were: tourist'sincome, tourism price, destination marketing, policies, and practices. Of these factors, tourist's income,tourism prices and exchange rates had the greatest impact on inbound tourism demand. This finding reveals that economic factors dominate the impact of tourism demand. Many scholars' studies confirm thisconclusion (Aki, 1998; Manuel Vanegas & Croes, 2000; Dwyer et al., 2002). With the deepening of research,an increasing number scholars have studied various economic indicators associated with tourism demand.Early measures used to study the demand for tourism included the following: tourism consumption/income, tourism export/import, destination stay time, accommodation facility stay time (Lim, 1997).However, since the 1990s, the major proxy variables that are used to study tourism demand are touristarrivings and tourism expenditures (Li et al., 2006). After the 21st century, the use of tourist volume asa proxy variable for international tourism demand has become increasingly common (Song & Li, 2008).Among all variables, the most suitable for measuring income is disposable income. However, this variableis difficult to obtain. Therefore, the nominal/real GDP, GNP or GDP per capita and GNP per capita aregenerally used instead. In a recent study, based on a sample of 11 major tourist destinations in the UnitedStates,Chi (2015) defined the notion of \world per capita GDP" (relative to the United States) as a revenuevariable. He used the weighted average of the U.S. dollar exchange rate against major source countries asthe nominal exchange rate variable. The study found that tourist demand is more sensitive to changes inincome than changes in exchange rates. Studies have also analyzed Industrial Production Index (IPI) as aproxy variable for tourism revenue, noting that IPI is not a good proxy for alternative travel revenue (Dogruet al., 2017). Price factors often use relative prices, exchange rates, transportation costs, opportunity costs,and the risk-spill prices of tourist destinations when selecting indicators (Crouch, 1994). Theoretically, thebest variable to measure the price of tourism products should be the tourist price index, but this indicator isvery difficult to measure. Therefore, most studies use the consumer price index instead. The relative priceis chosen as the result of the exchange rate adjustment based on the ratio of consumer price index (Lim, 1997). In addition, there are also studies that use consumerprice index, the hotel price index, or the result of a weighted average calculation of the consumer-relatedproduct/service price as the proxy variable of the travel price. Transportation costs are also importantfactors in tourism demand. However, due to the difficulty of estimating transportation costs, few studieshave included this factor in their models of tourism demand. The exchange rate factor is mainly used toadjust the price, however some studies have also used exchange rates as independent variables (Song & Li,2008).

When studying the demand for inbound tourism, many scholars also base their analysis on non-economicfactors and find different results. These studies mainly focus on the attractiveness and quality of service,political and government factors, social and cultural differences, and special events. Special events havean impact on the inbound tourism demand of a country (Loeb, 1982; Qu & Or, 2006). They can includeeconomic events (such as major sports events, financial crisis, etc.) as well as non-economic events (such asmajor diseases, social conflicts, terrorism, etc.). Therefore, in empirical models, special events are usuallytreated as dummy variables. In the study of non-economic factors, the attractiveness of the destinationand service quality are important factors to measure when examining the fluctuation in travel demand fortourist destinations. In summary, scholars mainly use quantitative methods to build models for the study of tourism demand.However, tourism demand is a complicated system, and the various factors affect each other. Because theresearch variables and research methods used to study tourism demand differ, the conclusions drawn willalso differ.

Tourism demand prediction method

Since the 1960s, predicting tourism demand has been a hot research topic. Scholars have devotedthemselves to the design of tourism demand prediction models and have made many achievements. It ispossible to classify tourism demand prediction methods into qualitative and quantitative categories. Thequalitative method, represented by the Delphi method, solicits the opinions of tourism professionals topredict demand. This method has proven to be of great value in studying the factors that cannot beeasily quantified (Shafer et al., 1974). Qualitative methods also include jury of executive opinion, subjectiveprobability assessment, and consumer intentions surveys (Tao & Ni, 2010). Qualitative methods are based onthe opinions of tourism experts and experienced practitioners, and form a comprehensive view of all parties'opinions as the basis for predicting tourism demands. This method relies on expert's knowledge, experience,and judgment in actual operation. For this reason, it is greatly influenced by subjective factors, and it isdifficult to accurately describe the factors quantitatively. Therefore, the use of qualitative methods is quitecontroversial. Currently, most studies of tourism demand prediction are based on quantitative researchmethods. Quantitative methods include the use of non-causal time series models, the causal econometricapproaches, and artificial intelligence (Song & Li, 2008).

The non-causal time series models depend on the sequence correlation of variables with tourism demand.Inbound tourist quantity is predicted by the correlation between lagged variables and current variables.Since time series models only require historical observations of variables, they are less expensive in terms ofdata collection and model estimation and have been widely used in tourism demand prediction over the past50 years (Peng et al., 2014). Time series models include simple auto-regressive and moving average (ARMA) models, autoregressive integrated moving average(ARIMA) models, exponential smoothing (ES) models, generalized autoregressive conditional heteroskedasticity (GARCH) models, and structural time series models (Goh & Law, 2002; Chan et al., 2005;Jackman & Greenidge, 2010). Among them, the ARMA model proposed by Box & Jenkins (1970) is themost widely used. More than two-thirds of the studies conducted since 2000 have employed various derivedmodels of the ARMA model, including the seasonal ARIMA (i.e., SARIMA) modelsand the fractional integration ARIMA (i.e., ARFIMA) models (Cho, 2001; Song & Li, 2008; du Preez &Witt, 2003; Goh & Law, 2002; Smeral & Wger, 2005; Gounopoulos et al., 2012).

The causal econometric model studies the causal relationship between tourism demand and it influencingfactors from an economic perspective. As they are well-suited to tourism demand modeling, modern econometric methods such as the autoregressive distributed lag (ARDL) model, error correction model (ECM),vector autoregressive (VAR) model, time varying parameter (TVP) models, almost ideal demand system(AIDS), and basic structural model (BSM) are widely used in tourism demand prediction Stučka (2002);Fourie & Santana-Gallego (2011); Kulendran & Wilson (2000); Lim & McAleer (2001). Compared with thetime series model, the econometric model better reflects the causal relationship between tourism demandvariables and its influencing factors from an economic point of view. This perspective can provide empiricalevidence for government tourism policy formulation and corporate strategic layout. Furthermore, becausetourism demand is affected by many factors, the time series model is quite limited and its predictive poweris relatively poor.

In addition to econometric models and time series models, new quantitative forecasting methods, such asartificial intelligence, are beginning to be applied to tourism demand forecasting. Methods include rough sets,genetic algorithms, ANN artificial neural networks and GFM gray theory prediction models. The rough setmethod can effectively analyze incomplete and inaccurate information and deduce the essential rules behindthis chaotic information. Because of this, it often used to enhance tourism demand forecasting (Law & Au,2000). Drawing from the theory of biological evolution, a genetic algorithm searches for the optimal solutionby simulating the natural evolutionary process.Hernández-López & Cáceres-Hernández (2007) confirmed thatgenetic algorithm can be applied to the prediction of tourism demand. Artificial neural network methodssimulate several human brain functions, and is capable of parallel processing, self-learning, self-organizingand has the ability to adapt. This ability enables this method to address several problems such as incompleteand non-linear tourism data information, numerous influencing factors, and large degrees of uncertainty.Because of these qualities, this method is able to compensate for the shortcomings of traditional forecastingmethods (Kon & Turner, 2005). The gray forecasting method is based on the gray theory system and isable to handle small sample size, lack of data, and greater uncertainty. Due to the significant instabilityand volatility of tourism development and the lack of data collection and statistics on tourism developmentin China, the gray model is widely used for the prediction of domestic tourism market. Based on previous studies, we anticipate that the tourism prediction model will show different degrees ofpredictive ability for different data modeling frequencies, forecast period lengths, sources, and destinations(Song & Witt, 2000). In order to improve the accuracy of tourism demand prediction, scholars no longerrely on a single forecasting model, but instead use a combination of forecasting models and then comparethe results to improve prediction accuracy (Law, 2000).

2.3 Forecasting with search trend data

Recently, internet search data has started to be been taken seriously by scholars to predict social economicactivities including stock market transactions, consumption levels, unemployment rates, and house priceswith a high degree of accuracy (Askitas & Zimmermann, 2009; Da et al., 2011; Vosen & Schmidt, 2011; Wu& Brynjolfsson, 2013). Google and Baidu have introduced Google Trends and Baidu Index to statisticallyanalyze the features of keywords submitted by search engine users. Scholars have studied the relationshipbetween the number of searches for a tourist destination and its actual market demand, and found that therewas a significant correlation between the two. In addition, the introduction of search keywords improvedthe accuracy of the prediction model. Choi & Varian (2012) used Google search engine data to predicttourism demand. Huang et al. (2013) used the Baidu Index to study the relationship between keywordsand the actual number of tourists visiting Beijing's Forbidden City. Their results provided the decision-making basis for the management of the Forbidden City Scenic Area by accurately predicting the number of tourists.Bangwayo-Skeete & Skeete (2015) combined Google Trends and autoregressive mixed data sampling modelsto improve the accuracy of visitor predictions. Yang et al. (2015) used internet search queries to predictthe number of tourists in Hainan, China and compared the predictive power of the search engine data fromGoogle and Baidu. The results showed that both sets of data significantly reduced the prediction error,however Baidu has a larger market share in China, and therefore the data performed better. As we know,Google search engine occupies more than Baidu in the world market, and can more fully reflect and predictthe flow of tourist information of inbound tourists. Here, we selected keywords from Google Trends andused the ARMA and ARDL model to measure the demand for tourism to China.

Research methodology and data collection

Prediction framework

First, we established a prediction framework for the number of inbound tourists in China. The research process was divided into four steps：

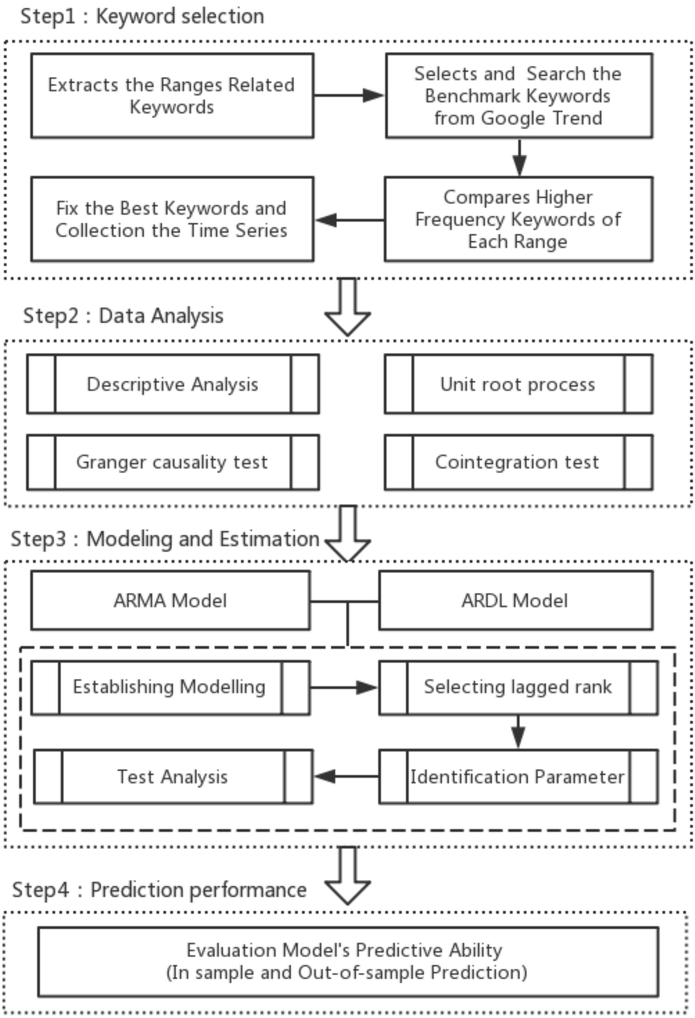


Fig 1. The framework of incorporating search information into tourism predictions

1) *Selection of keywords*. We identified four factors that influence the demand for tourism: the attractiveness of tourist destinations; economic factors; political factors; and cultural distance (Ritchie & Crouch, 2004). We then searched the benchmark keywords using Google Trends. We choose keywords with thehighest

frequencyfor each factor, and analyzed the correlation between the relative search frequency of keywords and the number of inbound tourists in China. We then selected the keywords with the highest correlation to use as the proxy variables.

2)*Data processing*. We validated the stability of the time series databy using unit root and cointegration tests to avoid encountering a false positive in our regression analysis. In addition, we used a Granger causality test to determine whether one variable had a lagged effect on the other and to examine the predictive power between variables.

3) *Establishment and analysis of the predictionmodel*. The main models used to test the causality of multivariate time series data include the ARMA, ARDL, and VAR models (Brooks,2008).Here, we chose to use the ARMA and ARDL models. We will use the ARMA model first. Then, we will add the Google search keywords as independent variables to the model in order to establish the ARDL model. We compare the predictions of these both models with the actual number of tourists, and evaluate the fitting effect of the prediction model.

4)*Accuracyevaluation of the prediction model*. The ARMA model and the ARDL model were used for in sample and out-of-sample predictions, and then logarithms were used to calculate the corresponding seasonally adjusted forecast values of actual tourist volume. We then compared the predictions of the two models to evaluate and predict their accuracy.The research framework is shown in Figure 1.

Keyword selection from Google Trends

The appropriate selection of keywords is fundamental for research on the correlation between search dataand economic behavior. However, there is no consensus on the best methods for selecting keywords. Atpresent, there are three main ways to choose keywords: technical word-taking method, direct word-takingmethod and range-based word-taking. Technical word-taking uses high-performance, large-scale computingequipment which takes all possible keywords, then compiles the relevant statistical model program to selectthe core keywords. Direct word-taking relies on subjective experience to determine the key words. Rangebased word-taking defines the range of a selected word, and then select the keywords in the range. Althoughthe accuracy of the technical word-taking method is highest, it is greatly restricted by the research conditionsand requires a large number of high-speed computers. The direct word-taking method and range-based wordtaking methods reduce the workload drastically, but there is a risk of missing the core keywords (R. Tierney& Pan, 2012; Kholodilin et al., 2009; DAmuri & Marcucci, 2010). In this paper, we chose to use the rangebased word-taking and direct word-taking methods. The selected four ranges evaluate tourist attractionsbased on destination attractiveness, economic factors, political factors, and cultural distance, respectively.

The first factor, destination attractiveness, reflects the feelings and opinions of its visitors about thedestination’s perceived ability to satisfy their needs (Vengesayi, 2003). The literature on destination attractiveness emphasizes that it is not only an objective summation of attractive elements, but also a system ofattraction, which is subordinate to a subsystem under the entire tourism system (Leiper, 1990). MacCannell(1976) proposed that the tourist attraction should include three components: a tourist, a sight, and a markeror image. However, the primary element of tourist attraction is its location (Leiper, 1990). Therefore, weused the search term "places in China" to identify benchmark key words. Based on this, the keywordscan be confirmed using the Google Trends database. The high-frequency keywords selected include tourismresources and tourism products, tourism facilities, hotel accommodations, foods and beverages, traffic conditions, and entertainment activities.

The second, economic factors, include incomes, relative prices, exchange rate,transportation costs, promotional expenditures, and international trade (Uysal & Crompton, 1984; Crouch, 1995; Li et al., 2005;Peng et al., 2015). In general, Income and Price were the most important factors. However, when searchingfor information, tourists focused on tourism expenditure, so here we use "cost in China" as the benchmarkkeyword.

Third, political factors, include wars, terrorism, and the state of international relations (Prideaux, 2005).Often, political factors arise that are beyond a country's ability to control. Since undergoing reform andopening up, China has provided visa-free policies to 51 countries and set up visa-issuing points at 87 ports.Select foreign tourists can enjoy 72-hour transit visa-free policies in over 10 cities[2]. Combined with thesearch habits of Google users, we selected benchmark keywords from several aspects of China's visa policy,China's customs tariff policy, and other tourism-related policies.

The fourth factor is cultural distance. The four main elements of culture we identified that were likelyto impact a tourist's destination choice were: the tourist's national culture; the tourist's individual level(internalized) culture; a destination'culture; and the "distance" between a tourist's home culture and adestination's culture (Ng et al., 2007). According to previous research, the difference between a tourists'psychological and cultural differences and those of the tourist destination is a draw for some tourists. Forthis reason, some foreign consumers are interested in the special culture of China. However, some touristsdo not seek out this cultural difference, and in turn are not interested in the tourism market of China.Therefore, it is unclear whether cultural distance facilitates the development of China's inbound tourism or,rather acts as an obstacle to development. Here, we use "Spoken Chinese" as the benchmark keyword, andthe high-frequency key words focus on Chinese etiquette, customs, language, diet, and architecture.

As described, we collected the time series data for high-frequency keywords in the four ranges fromGoogle Trends. Then we analyzed the correlation between these data and the time series data for Chineseinbound tourists. According to the results of the correlation analysis, the four key words with the highestrelative

[2]http://www.sohu.com/a/131220110_271706.

frequency and the largest correlation coefficient were selected. They were "Attractions in China","CNY", "China visa policy" and "Spoken Chinese" represented by AIC, CNY , CVP and SC.The correlation analysisresults are shown in Table 1.

Table 1 Key words selection and correlation coefficients

Key words	**Pearson coefficient**	**Key words**	**Pearson coefficient**	**Key words**	**Pearson coefficient**	**Key words**	**Pearson coefficient**
Places in china	0.103	Cost in China	-0.228**	China visa policy	0.239**	Spoken Chinese	0.209**
Attractions in China	0.185**	CNY exchange rate	0.075	Visa for China	0.238	Chinese culture	-0.170**
Hotels in China	0.147	Cheapest flight to China	-0.066	Tariffs on China	-0.128	Custom in China	-0.337**
great wall of China	0.06	China tickets	-0.059	Chinese visa	-0.069	Chinese dress code	-0.347**
Amazing places in china	0.081	Cheap China tours	0.186**	One belt one road	0.01	Chinese etiquette	-0.155**
Tourist attractions in china	0.184**	CNY	0.230**	New silk road	0.221**	Chinese language	-0.353**
Scenery of China	0.09	Budget to china	0.185**	Open policy in China	0.017	Chinese manners	-0.303
Chinese history museum	-0.97	Cheap flight to China	0.076			Chinese table manners	0.134**
Chinese museum	-0.017	China airlines economy class	0.119			Festivals in China	0.007
Chinese World Expo	-0.085	China ticket prices	0.006			Chinese values	0.012
Tiananmen	-0.065	Chinese buffet prices	0.182**			Regions of China	0.02
Shaolin temple	-0.15	Exchange rate USD to RMB	0.181**			Food in China	0.189**
The palace museum	-0.125	RMB conversion	0.079				
Terra cotta warriors	-0.243**	Transportation in China	-0.337**				
Potala palace	0.09	USD to RMB conversion	0.086				
Mount huangshan	0.125	USD to RMB exchange rate	0.081**				
Shopping in China	0.166	Trip to China cost	0.011				
Restaurant in China	0.048	RMB exchange rate	0.017**				

Notes: the symbols **denote the rejection of the null hypothesis at the 0.05 significance level, respectively.

Sample and data

This study used monthly time-series data on China's inbound tourist arrivals along with the Google searchdata for select keywords. The data of tourist arrivals in China was obtained from the official websiteof China National Tourism Administration[3]. This paper uses seasonal adjustment method to deal with thetime series firstly. Each keyword time series was obtained from Google Trends at http://trends.google.com/trends/.The range of our sample is from April 2005 to February 2016 (131 samples). After obtaining the researchdata, this paper used inbound tourist data and keyword search data from April 2005 to December 2014as in-sample data for the study of the correlation between Google Trends and inbound tourists time seriesdata. Inbound tourist data and keyword search data from January 2015 to February 2016 were used forout-sample prediction.

Analysis and approach

Unit root, Cointegration,and Granger causality tests

In order to ensure the stability of the time series data and avoid encountering a false positive in ourregression analysis, we used a unit root test and a cointegration test before the regression model wasestablished. The unit root test we used is based on the Augmented Dickey-Fuller (ADF) test. Resultsof the ADF test showed that the time series data of inbound tourists to China is nonstationary but thelogarithm series is stationary. To address this, all of the data sequences were processedusing a logarithmic transformation in order to ensure the unity of economic significance. The letter *L* infront of the variable names stands for logarithm. The test results are shown in Table 2.

Table 2 The results of the variables unit root test

	ADF statistic	P-values		ADF statistic	P-values
T	-2.356	0.157	CNY	-8.637	0.000
LT	-3.052	0.033	LCNY	-9.390	0.000
AIC	-5.851	0.000	SC	-10.983	0.000
LAIC	-5.859	0.000	LSC	-8.605	0.000
CVP	-9.760	0.000			
LCVP	-3.591	0.000			

Table 3. The results of Granger causality test

Null hypothesis	F-statistics	P-value
LAIC does not Granger Cause LT	0.105	0.957
LT does not Granger Cause LAIC	4.043	0.009
LCVP does not Granger Cause LT	1.164	0.317
LT does not Granger Cause LCVP	4.55	0.013
LCNY does not Granger Cause LT	2.032	0.08
LT does not Granger Cause LCNY	0.591	0.707
LSC does not Granger Cause LT	5.609	0.005
LT does not Granger Cause LSC	2.201	0.116

[3]China National Tourism Administration website:http://www.cnta.gov.cn/

Cointegration analysis was used to determine whether there is a long-term equilibrium between variablesrather than short-term fluctuations. We used two commonly used cointegration test methods, the EngelGranger two-step test and Johansen test (Engle & Granger, 1987). To examine cointegration between multivariate data, and we selected Johnsen test for cointegration test. The first step is to establish a VAR modelto measure the change of AIC and SC with the lag order *p*, and determine whether the lag order of themodel is 12th order (Shown in the Appendix A). The second step, is to conduct the Johansen cointegrationtest based on the VAR model. According to the results, at the 5% significance level, there were five instancesof cointegration (Shown in the Appendix B). This shows that the amount of inbound tourists and Googlekeywords cointegrates as a long-term equilibrium relationship. Therefore, the variables are co-integrated,allowing the regression model to be established without the concern of encountering a false positive in ourresults.We also observed a cointegrated relationship between the independent variable and the dependentvariable, suggesting that Granger causality may play a role. In order to improve the accuracy of our timevariable prediction, we tested the selected five sequences using a Granger causality test. Results are shownin Table 3.

According to the Granger causality test, the "*LAIC*" and "*LCV P*" variables have a one-way causalrelationship, leading with the "*LT*" variable. In other words, the "*LT*" variable is the Granger cause of the"*LAIC*" and "*LCV P*". The variables "*LCNY* " and "*LSC*" have a one-way causal relationship with thevariable "*LT* ". In other words, the "*LCNY* " and "*LSC*" variables are the Granger cause of the variable"*LT* ".

Predictionmodelling

This paper compares the performance of ARMA and ARDL models. In order to test the ability of Googlesearch data to predict the number of inbound tourists in China, first we designed an ARMA model basedon the actual number inbound tourists time series data and performed an in-sample prediction. Then, weadded Google search keywords to the model as an independent variable to establish the ARDL model andperform in-sample prediction. Next, we compared the prediction results of these two models. The ARMAmodel was based on the actual time series data of inbound tourists and contained one or more lag valuesof the dependent variable. The ARDL model included both the lagged value of the dependent variable andthe lagged value of the Google search keywords variable(Pesaran et al., 1995; Pesaran & Shin, 1996).

Assuming that LT_t is the linear function of the estimated value at time t as the sum of q terms that represents the average random variation over q previous periods (the MA component), plus the sum of p AR terms that compute the current value of LT as the weighted sum of the p most recent values. The general form of the ARMA model is:

$$LT_t = \mathrm{c} + \varepsilon_t + \sum_{i=0}^{p} \varphi_i \, LT_{t-i} + \sum_{j=0}^{q} \theta_j \varepsilon_{t-j} \qquad (1)$$

Where e_t is a white noise sequence with a mean of 0 and a variance of s^2. LT_t represents the number of tourist arrivals to China at time t. According to the results of correlation analysis, the autocorrelation coefficient of the sequence"LT" gradually approaches zero. The partial correlation coefficient of the first order and second order is beyond the two times the standard deviation, that is,the partial correlation function is truncated after the second order. Combined with autocorrelation and partial correlation diagram,we can initially determine the formulation of the model.

The main goal of our study is to explore whether tourists' internet search data can help predict thenumber of inbound tourists. Therefore, we choose ARDL model to explore it. As a method of examininglong term and cointegrating relationships between variables, the ARDL approach to cointegration will giverealistic and efficient estimates (Pesaran et al., 1999). Unlike the Johansen & Juselius (1990) cointegrationprocedure, the ARDL approach to cointegration helps in identifying the cointegrating vector(s). That is,each of the underlying variables stands as a single long run relationship equation (Nkoro et al., 2016). Takinginto account the "Google Trends" lag effect on the actual number of tourists, we added four sequences of"LAIC", "LCVP", "LCNY" and "LSC" and their different lag periods as independent variables into themodel. The ARDL(p,q) model specification is given as follows:

$$LT_t = \mu + \sum_{i=0}^{P_1} \phi_i \, LT_{t-i} + \sum_{i=0}^{P_2} \beta_{AIC,i} LAIC_{t-i} + \sum_{i=0}^{P_2} \beta_{CVP,i} LCVP_{t-i} + \sum_{i=0}^{P_3} \beta_{CNY,i} LCNY_{t-i} + \sum_{i=0}^{P_4} \beta_{SC,i} LSC_{t-i} + \varepsilon_t \qquad (2)$$

The definitions of the variables in the foregoing equation are the same as those in Equation (1).The errorterm, e_t, is assumed to be independently and identically distributed (i.i.d.). $LAIC_t$ represents the GoogleTrends time series data for "Attractions in China"at time t. $LCVP_t$ represents the Google Trends timeseries data for "China visa policy"at time t. $LCNY_t$ represents the Google Trends time series data for"CNY" at time t LSC_t represents the Google Trends time series data for "Spoken Chinese" at time t.

Analysis results

Identification of parameters

According to the selection process of used for the ARMA model,we calculated the significant P-valuesfor each of the models and compared them to the AIC values. The results show that: the AR(3) modelfits well. This model had the smallest P-value is the smallest, and the parameters have had a significantimpact on the explanatory variables at the 5% significance level. Then, using orthogonal least square (OLS) to estimate each ARDLmodel, the lag order is was determined according to AIC and SC criteria. After repeated screening andestimation of independent variables and different lag periods of dependent variables, we determined the finalformulation of the model. The estimation results of ARMA model and ARDL model are as described inTable 4.

Tab. 4 Estimation results using different econometric models

ARMA model				ARDL model			
Variables	Coefficients			Variables	Coefficients		
C	6.977***			C	0.687		
AR(1)	0.320***			LAIC(- 4)	0.022**		
AR(2)	0.393***			LCVP(-5)	0.017**		
AR(3)	0.188**			LCNY(-3)	0.021***		
SIGMASQ	0.001***			LSC(-1)	0.046***		
				LT(-1)	0.432***		
				LT(-2)	0.412***		
Adj-R^2	0.597			Adj-R2	0.622		
Log likelihood	265.945			Log likelihood	253.053		
AIC	-4.461			AIC	-4.599		
SC	-4.343			SC	-4.424		
DW	1.982			DW	1.926		
ADF test statistic		-11.286	0.000	ADF test statistic		-10.125	0.000
Test critical values:	1% level	-3.488	——	Test critical values:	1% level	-3.496	——
	5% level	-2.887	——		5% level	-2.89	——
	10% level	-2.58	——		10% level	-2.582	——

Notes: the symbols **,***denote the rejection of the null hypothesis at the 0.05 and 0.01 significance level, respectively.

According to the estimation results of the ARMA model, the model AIC value is -4.461, the SC value is -4.343, the coefficient of determination (R^2) is 0.611, and the adjustment coefficient ($Adj - R^2$) is 0.597, which shows that the model fits well and according to the variable LT , 59.7% of the variability can be explained by this model.According to the estimation results obtained using the ARDL model, we found: (1) the $R^2$is 0.644, the coefficient of adjustment ($Adj-R^2$) is 0.622, indicating that the fit of ARDL model is good, and it shows that 62.2% of LT variability can be explained. (2) The significant values of all of the explanatory variables were less than 10%. In the confidence interval, the explanatory variables are better explained. The relative frequency of the Google search of the keyword “Attractions in China”can explain 2.25% of the variation of inbound tourists, that is to say, for every additional 1 unit of searching for the keyword “Attractions in China”“ when the other factors remain unchanged,tourist volumewill increase by 2.25%. The relative frequency of the Google search of the keyword “China visa policy” can explain the variation of inbound tourists by 1.69%. The relative frequency of Google searches for keyword “CNY” can explain 2.12% of inbound tourist variation. The relative frequency of Google searches keyword “SC” can explain the variation of inbound tourist volume of 4.73%.

4.3.2Validation of Analysis

After establishing the ARMA and ARDL models, we needed to verify their stability. We generated amodel residual sequence, and observed the trend of the residual sequence (Appendix C), test whether theresidual sequence exists in the unit root (results shown in Table 4) to judge whether the residual sequenceexist in the autocorrelation. According to the residual trend of the ARMA, we can see that the residualsequence is stable and fluctuates in a straight line with a value of zero. According to the ARMA residualunit root test results, the residual ADF test value was -11.286 and -10.125, falling to the left of the 1%significance threshold, indicating that the residual sequence is stable at the 99% confidence level. If weassume that that the residual sequence is white noise and that there is no unit root, then both the ARMAmodel and ARDL models pass the stationary test.

More specifically, since the time series data of the keywords was added as an independent variable in theARDL model, it is necessary to test the heteroscedasticity of the model and the multicollinearity betweenthe independent variables (results shown in Tables 5 and 6).

Tab. 5Heteroscedasticity test of ARDL model

F- Statistic	0.971	P(6,100)-value	0.449
Obs*R-squared	5.891	P.Chi-Square value	0.436
Scaled explained SS	6.217	P.Chi-Square value	0.399

Tab. 6Multicollinearity test of ARDL model

	LT	LCNY	LAIC	LCVP	LSC
LT	1	0.483	0.095	-0.368	-0.204
LCNY	0.483	1	0.041	-0.288	-0.457
LAIC	0.095	0.041	1	-0.187	0.018
LCVP	-0.368	-0.288	-0.187	1	0.093

LSC	-0.204	-0.457	0.018	0.093	1

As can be seen from Table 5, both the statistical analysis and the accompanying probabilities are greaterthan 5% of the significance level. Therefore, the original hypothesis of heteroscedasticity in the acceptedmodel does not exist, and the model passes the heteroscedasticity test. Parameter estimation was obtainedby the model using an effective estimator. From Table 6, we can see that the correlation coefficients betweenthe two explanatory variables of the model are below 80%, indicating that there is no multicollinearity.

Prediction performance

After the two models were estimated, they can be used to predict the dependent variable *LT* in thesample (from April 2005 to December 2014). We can then compare the prediction results with the actualtourist sequence data, *LT*. As shown in Figure 2, the predicted results of the ARMA and ARDL modelsare conservative compared with the actual values, and the general trend is the same. The fitting results ofthe two models are all ideal.The empirical results show that the root mean square error of the prediction results of the ARDL is0.023, and the root mean square of the ARMA is 0.024, indicating that the prediction of ARDL model ismore accurate.

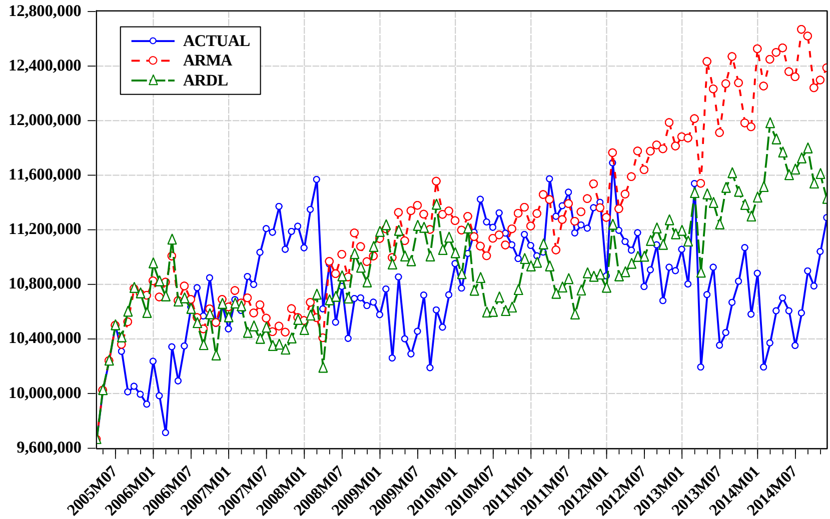


Fig 2. Comparison sample prediction using different models

To fully compare the predictive ability of the two models, we can use the models to predict the *LT*values of the explanatory variables outside the sample from January 2015 to February 2016. Then by usinga logarithmic transformation, we can calculate the corresponding seasonally adjusted forecast of the actualtourist volume, and compare the two models' predictions. At the time of the prediction, the two models arepredicting a month forward, substituting the prediction results into the model, and making the predictionfor the next month. The comparison between the two model predictions is as show in Table 3.

Here, we use the root mean square error index to compare the prediction accuracy of the two models.The root mean square error of the predicted value can be calculated by using the error of the predictedvalues of the two models. The results showed that the root mean square error of the ARMA predictivevalue was 19.188, and the root mean square error of the ARDL predictive value was 16.349. Therefore, theprediction accuracy of ARDL was higher than that of ARMA by 14.80%.

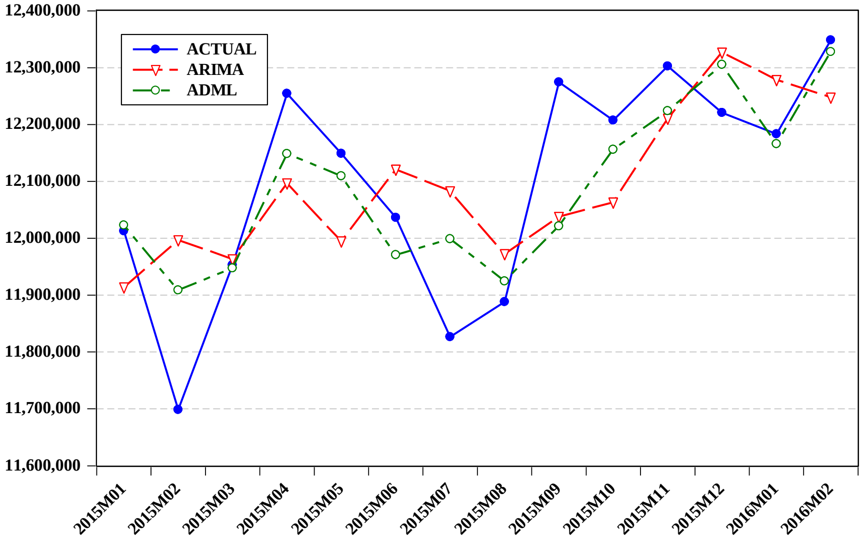


Fig 3. Out-of-sample prediction of the different models

The following is a summary of the empirical process and prediction comparison. Here we constructedtwo models to predict the inbound tourists. The two models were: the ARDL model with four Google searchindex data as independent variables and the ARMA without independent variables. Table 7 shows the fitof the ARMA model and predictive statistics comparisons of the two models with or without independentvariables.

Table 7 Prediction statistics of the different models

Fitting statistics	ARMA	ARDL
Adj-R^2	0.597	0.622
AIC	-4.461	-4.599
RMS prediction error in sample	0.024	0.023
RMS prediction error out of sample	19.188	16.349

We found that the ARDL model had a larger coefficient of determination than the ARMA model, witha lower AIC criterion, and that the model fit was better. The root-mean-square error of the internal andexternal predictive values of ARDL model samples were smaller than that of the ARMA, and the predictionaccuracy was increased by 14.80%, indicating that the Google search data improves the prediction abilityof the time series predictive model.

In addition, according to the parameter estimation of ARDL, each parameter had a significant impacton the explanatory variables. That is, the relative frequency of Google search terms for the keywords“Attractions in China”, “CNY”, “China visa policy” and “spoken Chinese” all had a significant impacton the number of inbound tourists in China. According to the coefficients of each parameter, we observedthat the relative frequency of the keyword “Spoken Chinese”, which had the most influence on the numberof inbound tourists, was followed by the keyword “Attractions in China” and the keyword “CNY”. Thekeyword “China visa policy” had the least influence.

Discussion and conclusions

In this paper, we used the data describing inbound tourists from April 2005 to February 2016 in China asthe research sample and used the relative frequency of searches for four Google keywords as the explanatoryvariables. We constructed an ARDL model with the relative frequency of Google keyword search as theindependent variable and constructed an ARMA model without an independent variable. We used a datasample from April 2005 to December 2014 as test data to determine the degree to which the models fit thedata and evaluate their prediction ability. We the compared the root mean square error between the twomodels for predicting the number of inbound tourists from January 2015 to February 2016 in China. Theresearch draws the following main conclusions:

First, based on the results of previous studies, this paper categorizes the influences for inbound tourismdemand based on the following factors: attractiveness of tourism destination, economic factors, politicalfactors and cultural factors. We selected several relevant keywords to represent each factor (“Attractionsin China”, “CNY”, “China visa policy”, “Spoken Chinese”), analyzed search trends of Google queries,analyzed the relative frequency of inbound tourists and the keyword search. The result of Granger causalitytest showed that there was a significant one-way Granger causality between the seasonally adjusted andlogarithmically adjusted inbound tourists and the log-transformed four Google keyword variables. Thisindicates that Google keyword variables help to explain the future changes in actual visitor numbers. Theempirical results of the ARDL model using the relative frequency of Google keyword search shows that thereis a long-term equilibrium relationship and positive correlation between the number of Chinese inboundtourists and queries for these four Google search keywords. In other words, as each Google keyword searchindex is added, the number of inbound tourists in China will also increase accordingly.

Second, by comparing the degree of fit for each model and the prediction accuracy for the data sample,we found that the ARDL model has a better fits and better reflects the changing trends of the explanatoryvariables. The ARDL model can also be used to determine the lagged model more accurately. By comparingthe prediction accuracy of the samples, we found that the root mean square error of the prediction of theARDL model is smaller and the prediction accuracy is higher, indicating that the relative frequency ofGoogle keyword search improves its prediction ability of the traditional time series model. The research ofHuang et al. (2013) and Yang et al. (2015) also proves that the introduction of search keywords can improvethe prediction accuracy of the traditional prediction model. Google Trends reports nearly 1 hour of keywordsearch relative frequency data. According to the ARDL model, we know that we can predict the relativeamount of inbound tourists relatively accurately and in a timely fashion as long as we know the relativefrequencies of the keywords “Attractions in China”, “China visa policy” and “CNY” are 4, 5, 3, and “SpokenChinese” is 1 month beforehand. These predictions can provide the relevant management departments withinformation that is critical when decision-making.

Third, our results show that the relative frequency of Google searches for China’s tourist attractionshas a significant impact on the changes in the number of inbound tourists. However, the coefficient of itsparameters is relatively small, indicating that some potential tourists find alternative tourism destinationsin China when obtaining travel information through the Internet. In this regard, the destination shouldtake into consideration the great potential of the Internet in overseas tourism promotion, and make activeuse of the Internet for tourism promotion activities to attract international tourists. In order to enhance theattractiveness and competitiveness of Chinese tourist destinations, the destinations should take into accounthow tourists from different countries and regions understand China differently. For example, Americantourists like Chinese monuments, Japanese tourists like Chinese food, Southeast Asian tourists are mostinterested in China’s landscape, history, and culture. According to tourists’ different perceptions of China,tourist destinations can integrate elements that appeal to a particular groups of tourist’s interests. Theycan then customize tourist information websites to target tourists from different markets, and facilitate theirexploration of Chinese culture.

Fourth, our results demonstrate that although the Google search volume regarding visa policies has asignificant impact on the volume of inbound tourists, the coefficient of its parameters is the smallest relativeto other keywords. This indicates that China’s visa policy needs to be further improved to encourage tourismdemand (Song et al., 2012). In order to promote the development of inbound tourism, it is necessary tofacilitate overseas travel for tourists to China. In 2015, the State Council of the People’s Republic of Chinaissued the “Opinions of the State Council on Promoting the Reform and Development of Tourism Industry.”It proposed to study how to facilitate the entry of foreigners into inbound tourism visas and to promoteforeigners visa services to foreigners at qualified ports of entry and to gradually optimize and perfect the72-hour transit visa-free policy for foreigners.The government can formulate special preferential visa policiesfor specific tourist markets. On the one hand, it can promote the development of regional economy in ourcountry. On the other hand, the government can increase the tourists' satisfaction with China's tourismand achieve the sustainable development of inbound tourism (Neiman & Swagel, 2009).

Fifth, the relative frequency of exchange rates using a Google search is also a significant factor affectingthe volume of inbound tourists in China. Exchange rates are often used as an important factor in the choiceof destination and its changes have a significant impact on inbound tourism (Muchapondwa & Pimhidzai,2011). Both when the RMB exchange rate rises, and correspondingly, the prices of Chinese tourism productsdenominated in RMB rise, or when the exchange rate of RMB declines, and the prices of Chinese tourismproducts drop, changes in the prices of tourism products will directly affect the volume of inbound tourists(Vita et al., 2013). Travel policy makers should closely monitor changes in exchange rates and promptlypropose appropriate incentives to develop tourism. On the one hand, tourism destinations should standardizethe assignment of prices for tourism products and services, and adjust the corresponding countermeasuresaccording to exchange rate changes (Kim & Lee, 2017). On the other hand, relying on a price advantage toattract more international tourists will be influenced by RMB appreciation and depreciation. Therefore, thisapproach should be adjusted to win over the tourism market with the quality of the products and services.Travel agencies can set up specialized overseas tourism marketing agencies to enhance the capability ofindependent outbound tourist teams and reduce the negative impact of RMB exchange rate changes.

Finally, the relative frequency of Chinese-speaking Google searches is also a significant factor affectingthe volume of inbound tourists in China. Our results show that tourists will search for "Chinese spoken"using Google one month prior to their visit to China for the purpose of learning Chinese. To this end,travel agencies, in cooperation with Chinese training institutions, can set up Chinese language learningwebsites specially designed for foreign tourists. On the one hand, they can shorten the cultural distance forforeign tourists and on the other hand, they can enhance the popularity of travel agencies and attract moreinternational tourists.

Appendix A

Table 8 The results of Lag order selection criteria

Lag	LogL	LR	FPE	AIC	SC	HQ
0	160.522	——	0	-3.888	-3.739	-3.828
1	234.996	137.778	0	-5.125	-4.232*	-4.766*
2	271.809	63.502	0	-5.42	-3.783	-4.764
3	303.444	50.617	0	-5.586	-3.204	-4.631

4	321.836	27.128	0	-5.421	-2.294	-4.167
5	345.109	31.419	0	-5.378	-1.507	-3.826
6	373.655	34.969	0	-5.466	-0.851	-3.616
7	393.208	21.508	0	-5.33	0.029	-3.181
8	418.01	24.182	0	-5.325	0.779	-2.878
9	460.306	35.952	0	-5.758	1.091	-3.012
10	521.283	44.208*	0	-6.657	0.936	-3.613
11	567.938	27.993	0	-7.198	1.139	-3.856
12	637.587	33.083	0	-8.315*a	0.767	-4.674

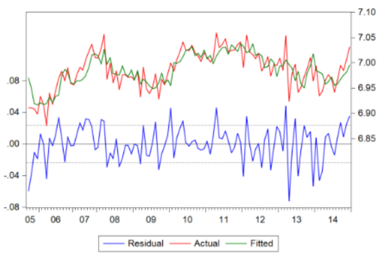
Note: [a] indicates optimal lag order selected by the criterion.

Appendix B

Table 9 The results of Lag order selection criteria

No.of CE(s)	Eigenvalue	Statistic	Critical Value	Prob.**
None*	0.486	114.44	69.819	0
At most 1*	0.328	61.147	47.856	0.002
At most 2*	0.232	29.361	29.797	0.026
At most 3*	0.059	8.192	15.495	0.035
At most 4*	0.041	3.31	3.841	0.049

Appendix C



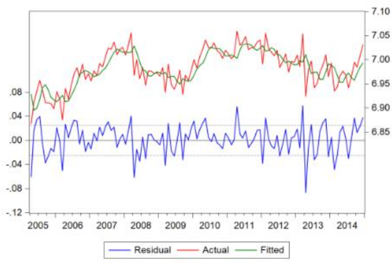


Fig. 4. Residual Sequence of ARMA model **Fig. 5. Residual Sequence of ARMA model**

References

Akis, S. (1998). A compact econometric model of tourism demand for Turkey. *Tourism Management*, *19*(1), 99-102.

Askitas, N., & Zimmermann, K. F. (2009). Google econometrics and unemployment forecasting. *Applied Economics Quarterly*, *55*(2), 107-120.

Bangwayo-Skeete, P. F., & Skeete, R. W. (2015). Can Googledata improve the forecasting performance of tourist arrivals? Mixed-data sampling approach. Tourism Management , 46, 454{464. URL: http://www.sciencedirect.com/science/article/pii/S0261517714001460. doi:10.1016/j.tourman.2014.07.014.

Botha, C., Crompton, J. L., & Kim, S.-S. (1999). Developinga Revised Competitive Position for Sun/Lost City, SouthAfrica. Journal of Travel Research, 37, 341-352. URL: http://journals.sagepub.com/doi/10.1177/004728759903700404.doi:10.1177/004728759903700404.

Box, G. E., & Jenkins, G. M. (1970). Time series analysis: forecasting and control. Oakland, California: Holden-Day.

Brooks, C. (2008). Introductory Econometrics for Finance, 2nd edition. Cambridge University Press.

Chan, F., Lim, C., & Mcaleer, M. (2005). Modelling multivariate internationaltourism demand and volatility. Tourism Management,26, 459-471.

Chi, J. (2015). Dynamic Impacts of Income and the Exchange Rateon US Tourism, 1960{2011. Tourism Economics, 21, 1047{1060.URL: https://doi.org/10.5367/te.2014.0399. doi:10.5367/te.2014.0399.China National Tourism Administration(CNTA) (2016). Chinatourism statistics bulletin in 2015. Avilable online at http://www.cnta.gov.cn/zwgk/lysj/201610/t20161018_786774.shtml.

Cho, V. (2001). Tourism forecasting and its relationshipwith leading economic indicators. Journal of Hospitality& Tourism Research, 25, 399{420. URL: https://doi.org/10.1177/109634800102500404.doi:10.1177/109634800102500404.arXiv:https://doi.org/10.1177/109634800102500404.

Choi, H., & Varian, H. (2012). Predicting the Present with GoogleTrends. Economic Record, 88, 2-9.

Crouch, G. I. (1994). Price Elasticities in International Tourism.Hospitality Research Journal, 17, 27-39. URL: https://doi.org/10.1177/109634809401700304. doi:10.1177/109634809401700304.

Crouch, G. I. (1995). A meta-analysis of tourism demand. Annalsof tourism research, 22, 103-118.

Da, Z., Engelberg, J., & Gao, P. (2011). In Search of Attention.The Journal of Finance, 66, 1461{1499. URL: http://www.jstor.org/stable/41305167.DAmuri, F., & Marcucci, J. (2010). Working Papers 2010.31 FondazioneEni Enrico Mattei. URL: https://EconPapers.repec.org/RePEc:fem:femwpa:2010.31.

Dann, G. M. S. (1977). Anomie, ego-enhancement and tourism.Annals of Tourism Research, 4, 184-194.

Dogru, T., Sirakaya-Turk, E., & Crouch, G. I. (2017). Remodelinginternational tourism demand: Old theory and newevidence. Tourism Management, 60, 47{55. URL: http://linkinghub.elsevier.com/retrieve/pii/S0261517716302114.doi:10.1016/j.tourman.2016.11.010.

Dwyer, L., Forsyth, P., & Rao, P. (2002). Destination price competitiveness:exchange rate changes versus domestic ination. Journalof Travel Research, 40, 328{336.

Engle, R., & Granger, C. (1987). Co-integration and Error Correction:Representation, Estimation, and Testing. Econometrica, 55, 251-76. URL: https://econpapers.repec.org/article/ecmemetrp/v_3a55_3ay_3a1987_3ai_3a2_3ap_3a251-76.htm.

Fourie, J., & Santana-Gallego, M. (2011). The impactof mega-sport events on tourist arrivals. Tourism Management, 32, 1364 { 1370. URL: http://www.sciencedirect.com/science/article/pii/S0261517711000148.doi:https://doi.org/10.1016/j.tourman.2011.01.011.

Frechtling, D. C. (1996). Transport for tourism : Stephen page.routledge. Annals of Tourism Research, 23, 727-729.

Goh, C., & Law, R. (2002). Modeling and forecasting tourism demandfor arrivals with stochastic nonstationary seasonality andintervention. Tourism Management, 23, 499-510.

Gounopoulos, D., Petmezas, D., & Santamaria, D. (2012).Forecasting Tourist Arrivals in Greece and the Impact ofMacroeconomic Shocks from the Countries of Tourists' Origin.Annals of Tourism Research, 39, 641-666. URL: http://www.sciencedirect.com/science/article/pii/S0160738311001496.doi:10.1016/j.annals.2011.09.001.

Guthrie, H. W. (1961). Demand for tourist' goods and services ina world market. Papers of the Regional Science Association, 7,159-175.

Hern_andez-L_opez, M., & C_aceres-Hern_andez, J. J. (2007). Forecastingtourists' characteristics by a genetic algorithm witha transition matrix. Tourism Management, 28, 290{297. URL: http://www.sciencedirect.com/science/article/pii/S026151770600029X.doi:10.1016/j.tourman.2005.11.016.

Huang, X., Zhang, L., & Ding, Y. (2013). Study onthe predictive and relationship between tourist attractionsand the baidu index|A case study of the forbiddencity. Tourism Tribune, 28, 93{100. URL: http://www.lyxk.com.cn/CN/abstract/abstract13597.shtml. doi:10.3969/j.issn.1002-5006.2013.011.011.

Jackman, M., & Greenidge, K. (2010). Modelling and forecastingtourist ows to barbados using structural time seriesmodels. Tourism and Hospitality Research, 10, 1{13. URL:https://doi.org/10.1057/thr.2009.23. doi:10.1057/thr.2009.23. arXiv:https://doi.org/10.1057/thr.2009.23.

Jeong, S. O., & Park, S. H. (1997). A cross-cultural application ofthe novelty scale. Annals of Tourism Research, 24, 238{240.

Johansen, S., & Juselius, K. (1990). Maximum likelihood estimationand inference on cointegrationwith applications to the demand formoney. Oxford Bulletin of Economics and statistics, 52, 169-210.

Johansen, S., (1988). Statistical analysis of cointegration vectors.Journal of Economic Dynamics and Control, 12, 231-254.

Kholodilin, K., Podstawski, M., Siliverstovs, B., & B• urgi, C. (2009).Google Searches as a Means of Improving the Nowcasts of KeyMacroeconomic Variables. DOI: 10.2139/ssrn.1507084.

Kim, H., & Richardson, S. L. (2003). Motion picture impacts ondestination images. Annals of Tourism Research, 30, 216{237.

Kim, J., & Lee, C.-K. (2017). Role of tourism price in attractinginternational tourists: The case of japanese inbound tourism fromsouth korea. Journal of Destination Marketing & Management,6, 76-83.

Kon, S. C., & Turner, L. W. (2005). Neural networkforecasting of tourism demand. Tourism Economics, 11, 301-328. URL: https://doi.org/10.5367/000000005774353006. doi:10.5367/000000005774353006.arXiv:https://doi.org/10.5367/000000005774353006.

Kozak, M. (2002). Destination benchmarking. Annals of tourismresearch, 29, 497-519.

Kulendran, N., & Wilson, K. (2000). Modelling business travel.Tourism Economics, 6, 47{59. URL: https://doi.org/10.5367/000000000101297460. doi:10.5367/000000000101297460.arXiv:https://doi.org/10.5367/000000000101297460.

Law, R. (2000). Back-propagation learning in improving theaccuracy of neural network-based tourism demand forecasting.Tourism Management, 21, 331{340. URL: http://www.sciencedirect.com/science/article/pii/S0261517799000679.doi:10.1016/S0261-5177(99)00067-9.

Law, R., & Au, N. (2000). Relationship modeling in tourism shopping:a decision rules induction approach. Tourism Management, 21, 241{249. URL: http://www.sciencedirect.com/science/article/pii/S0261517799000564. doi:10.1016/S0261-5177(99)00056-4.

Leiper, N. (1990). Tourist attraction systems. Annals of Tourism Research, 17, 367{384. URL: http://www.sciencedirect.com/science/article/pii/016073839090004B.doi:10.1016/0160-7383(90)90004-B.

Li, G., Song, H., & Witt, S. F. (2005). Recent Developments inEconometric Modeling and Forecasting. Journal of Travel Research, 44, 82{99. URL: http://journals.sagepub.com/doi/10.1177/0047287505276594. doi:10.1177/0047287505276594.

Li, G., Song, H., & Witt, S. F. (2006). Forecasting tourism demandusing econometric models. Tourism Management Dynamics, (pp.219-228).

Li, X., Pan, B., Law, R., & Huang, X. (2017). Forecasting tourismdemand with composite search index. Tourism Management, 59, 57{66. URL: http://linkinghub.elsevier.com/retrieve/pii/S0261517716301133.doi:10.1016/j.tourman.2016.07.005.

Lim, C. (1997). Review of international tourism demand models.Annals of Tourism Research, 24, 835-849.

Lim, C. (1999). A meta-analytic review of international tourismdemand. Journal of Travel Research, 37, 273-284.

Lim, C., & McAleer, M. (2001). Cointegration analysis of quarterlytourism demand by hong kong and singapore for australia.Applied Economics, 33, 1599{1619. URL: https://doi.org/10.1080/00036840010014012. doi:10.1080/00036840010014012.arXiv:https://doi.org/10.1080/00036840010014012.

Loeb, P. D. (1982). International travel to the United States:An econometric evaluation. Annals of Tourism Research, 9, 7{20. URL: http://www.sciencedirect.com/science/article/pii/0160738382900317. doi:https://doi.org/10.1016/0160-7383(82)90031-7.

Ma, Y., & Li, T. (1999). Research on inbound tourism in China.Science Press.

MacCannell, D. (1976). The tourist: A new theory of the leisureclass. Univ of California Press.

Manuel Vanegas, S., & Croes, R. R. (2000). Evaluation of demand :Us tourists to aruba. Annals of Tourism Research, 27, 946-963.

Muchapondwa, E., & Pimhidzai, O. (2011). Modelling internationaltourism demand for zimbabwe. International journal of businessand social science, 2.

Neiman, B., & Swagel, P. (2009). The impact of post-9/11 visapolicies on travel to the united states. Journal of InternationalEconomics, 78, 86-99.

Ng, S. I., Lee, J. A., & Soutar, G. N. (2007). Tourists intentionto visit a country: The impact of cultural distance.Tourism Management, 28, 1497{1506. URL: http://www.sciencedirect.com/science/article/pii/S0261517706002044.doi:10.1016/j.tourman.2006.11.005.

Nkoro, E., Uko, A. K. et al. (2016). Autoregressive distributedlag (ardl) cointegration technique: application and interpretation.Journal of Statistical and Econometric Methods, 5, 63-91.

Oh, H. C., Uysal, M., & Weaver, P. A. (1995). Product bundlesand market segments based on travel motivations: a canonicalcorrelation approach. International Journal of Hospitality Management, 14, 123{137. URL: http://www.cabdirect.org/abstracts/19961800960.html.

Paul C. Fakeye, & John L. Crompton (1991). Image Di_erences betweenProspective, First-Time, and Repeat Visitors to the LowerRio Grande Valley. Journal of Travel Research, 30, 10-16.

URL: https://doi.org/10.1177/004728759103000202. doi:10.1177/004728759103000202.

Peng, B., Song, H., & Crouch, G. I. (2014). A meta| analysis ofinternational tourism demand forecasting and implications forpractice. Tourism Management, 45, 181{193. URL: http://www.sciencedirect.com/science/article/pii/S0261517714000806.doi:https://doi.org/10.1016/j.tourman.2014.04.005.

Peng, B., Song, H., Crouch, G. I., & Witt, S. F. (2015). A metaanalysisof international tourism demand elasticities. Journal ofTravel Research, 54, 611-633.

Pesaran, H., Shin, Y. et al. (1995). Long-run structural modelling.Technical Report Faculty of Economics, University of Cambridge.

Pesaran, M. H., & Shin, Y. (1996). Cointegration and speed ofconvergence to equilibrium. Journal of econometrics, 71, 117-143.

Pesaran, M. H., Shin, Y., & Smith, R. P. (1999). Pooled meangroup estimation of dynamic heterogeneous panels. Journal ofthe American Statistical Association, 94, 621-634.

du Preez, J., & Witt, S. F. (2003). Univariate versusmultivariate time series forecasting: an applicationto international tourism demand. International Journal of Forecasting, 19, 435-451. URL:http://www.sciencedirect.com/science/article/pii/S0169207002000572.doi:https://doi.org/10.1016/S0169-2070(02)00057-2.

Prideaux, B. (2005). Factors a_ecting bilateral tourism ows.Annals of Tourism Research, 32, 780-801. URL: http://linkinghub.elsevier.com/retrieve/pii/S016073830500068X.doi:10.1016/j.annals.2004.04.008.

Qu, H., & Or, Y.-S. (2006). Determinants of the Travel DemandModel for Canadian Tourists to the U.S. International Journalof Hospitality & Tourism Administration, 7, 1-19.

R. Tierney, H. L., & Pan, B. (2012). A poisson regression examinationof the relationship between website tra_c and search enginequeries. NETNOMICS: Economic Research and Electronic

Networking, 13, 155-189. URL: https://doi.org/10.1007/s11066-013-9072-x. doi:10.1007/s11066-013-9072-x.

Ritchie, J. R. B., & Crouch, G. I. (2003). The Competitive Destination: A Sustainable Tourism Perspective. CABI. Google-Books-ID: yvydAwAAQBAJ.

Shafer, E. L., Moeller, G. H., & Getty, R. E. (1974). Future leisureenvironments. Ekistics-The Problems And Science Of Human Settlements, 40, 68-72.

Smeral, E., & Wger, M. (2005). Does complexity matter? Methodsfor improving forecasting accuracy in tourism: The case of austria.Journal of Travel Research, 44, 100{110. URL: https://doi.org/10.1177/0047287505276596. doi:10.1177/0047287505276596.arXiv:https://doi.org/10.1177/0047287505276596.

Song, H., Dwyer, L., Li, G., & Cao, Z. (2012). Tourism economicsresearch: A review and assessment. Annals of Tourism Research,39, 1653-1682.

Song, H., & Li, G. (2008). Tourism demand modelling and forecasting| A review of recent research. Tourism Management,29, 203-220. URL: http://linkinghub.elsevier.com/retrieve/pii/S0261517707001707.doi:10.1016/j.tourman.2007.07.016.

Song, H., & Witt, S. (2000). Tourism Demand Modelling and Forecasting. London: Routledge.

Stu_cka, T. (2002). A Comparison of Two Econometric Models (OLSand SUR) for Forecasting Croatian Tourism Arrivals. WorkingPapers 8 The Croatian National Bank, Croatia. URL: https://EconPapers.repec.org/RePEc:hnb:wpaper:8.

Tao, W., & Ni, M. (2010). Study on the comparison of tourism demandforecast between china and western countries: Basic theoryand models. Tourism Tribune, 25, 12-17.

Uysal, M., & Crompton, J. L. (1984). Determinants of demand for international tourist ows to Turkey. Tourism Management, 5,288-297.

Vengesayi, S. (2003). A conceptual model of tourism destinationcompetitiveness and attractiveness. (pp. 637-647).

Vita, G. d., Kyaw, K. S. et al. (2013). Role of the exchange rate intourism demand. Annals of Tourism Research, 43, 624-627.

Vosen, S., & Schmidt, T. (2011). Forecasting private consumption:survey-based indicators vs. Google trends. Journal of Forecasting,30, 565{578.World Tourism Orgnaization(UNWTO) (2016). Annual report 2016.Avilable online at http://cf.cdn.unwto.org/sites/all/files/pdf/annual_report_2016_web_0.pdf.

Wu, L., & Brynjolfsson, E. (2013). The future of prediction: Howgoogle searches foreshadow housing prices and sales. Social Science Electronic Publishing, (p. 147).

Yang, X., Pan, B., Evans, J. A., & Lv, B. (2015). Forecasting Chinesetourist volume with search engine data. Tourism Management, 46, 386-397. URL: http://www.sciencedirect.com/science/article/pii/S0261517714001514.doi:10.1016/j.tourman.2014.07.019.

Zheng, P., MA, Y., Wang, J., Bai, K., & Wang, X. (2014). Ananalysis of push and pull relationships in the foreign tourist toChina. Human Geography, 135, 146-153.

ABOUT THE AUTHORS

Huiyue Liu, Ph.D., assistant Professor at Shenzhen Tourism College of Jinan University, China. Her research interests are big data analytics, econometric modeling, macro economic decision analysis. E-mail:liu_hy@sz.jnu.edu.cn.

Gege Zhang, master student at Shenzhen Tourism college of Jinan University, China. Her research interests include econometric modeling, data analyzing and tourism forecasting. E-mail: zzhang_gg@126.com.