Tourism Management 59 (2017) 57–66

Contents lists available at ScienceDirect

Tourism Management

journal homepage: www.elsevier.com/locate/tourman

Forecasting tourism demand with composite search index

Xin Li [a,*], Bing Pan [b,c], Rob Law [d], Xiankai Huang [e]

[a] Institute of Tourism, Beijing Union University, Beijing, 100101, China
[b] Department of Recreation, Park, and Tourism, College of Health and Human Development, Penn State University, University Park, PA, 16802, USA
[c] School of Tourism and Environmental Sciences, Shaanxi Normal University, Xi'an, China
[d] School of Hotel and Tourism Management, The Hong Kong Polytechnic University, Kowloon, Hong Kong
[e] Beijing Open University, Beijing, 100081, China

HIGHLIGHTS

- We propose a framework to accurately forecast Chinese tourism demand.
- Search engine query data is collected to forecast tourist volumes to Beijing.
- A generalized dynamic factor model is used to create a composite search index.
- Our method improves forecast accuracy better than two benchmark models.

ARTICLE INFO

Article history:
Received 14 January 2016
Received in revised form
8 July 2016
Accepted 11 July 2016
Available online 29 July 2016

Keywords:
Tourism demand forecast
Big data analytics
Search query data
Generalized dynamic factor model
Composite search index

ABSTRACT

Researchers have adopted online data such as search engine query volumes to forecast tourism demand for a destination, including tourist numbers and hotel occupancy. However, the massive yet highly correlated query data pose challenges when researchers attempt to include them in the forecasting model. We propose a framework and procedure for creating a composite search index adopted in a generalized dynamic factor model (GDFM). This research empirically tests the framework in predicting tourist volumes to Beijing. Findings suggest that the proposed method improves the forecast accuracy better than two benchmark models: a traditional time series model and a model with an index created by principal component analysis. The method demonstrates the validity of the combination of composite search index and a GDFM.

1. Introduction

Advances in information technology have given rise to a massive amount of big data generated by users, including search queries, social media mentions, and mobile device locations (Mayer-Schonberger & Cukier, 2013). In particular, search query data provide valuable information about tourists' interests, opinions, and intentions. Tourists use search engines to obtain weather and traffic information, and to plan their routes by searching for hotels, attractions, travel guides, and other tourists' opinions (Fesenmaier, Xiang, Pan, & Law, 2011). Search query data, including its content and volume, can capture tourists' attention to travel destinations and can be useful in accurately forecasting tourist volumes. The abundant search trend data became a favorable source for tourism forecasting in the era of Big Data (Pan, Wu, & Song, 2012; Yang, Pan, & Song, 2014; Yang, Pan, Evans, & Lv, 2015). However, they also bring challenges in the modeling process of tourism forecasting.

In particular, in forecasting tourist volumes with search trend data, one needs to collect tourism-related keywords, obtain their search trend data, select appropriate data series to construct an aggregated index, and construct econometric models. The major challenges are keyword selection and search data aggregation. Keyword selection has received significant attention from researchers. For example, Brynjolfsson, Geva, and Reichman (2016) proposed a crowd-squared method. They prompted individuals through an online interface to produce word associations and the results verified that this method performed efficiently in the keyword selection task. In comparison, the process of index

* Corresponding author.
E-mail addresses: leexin111@163.com (X. Li), bingpan@gmail.com (B. Pan), rob.law@polyu.edu.hk (R. Law), xiankai@buu.edu.cn (X. Huang).

http://dx.doi.org/10.1016/j.tourman.2016.07.005
0261-5177/

aggregation has received limited attention (Brynjolfsson et al., 2016). This step is generally conducted through three main approaches: (1) incorporating keywords directly into the models; (2) extracting the index using principal component analysis (PCA); and (3) index aggregation from multiple variables (Yang et al., 2015). Although these approaches could predict more accurately than their benchmark models, they are still not optimal. First, high dimensions of variables may incur multicollinearity or overfitting problems (Varian, 2014). In particular, out-of-sample forecasts may fail even when in-sample forecasts perform well; second, a large amount of the original information will be lost if the data series is weighed equally in aggregating an index from multiple keywords. Incomplete information may reduce the forecast accuracy.

This study aims to propose a feasible variable selection method in forecasting tourist volumes with search trend data. In order to generate a universally acceptable framework, the approach should follow two rules: first, it should acquire one representative and meaningful index that reflects the dynamic correlation among all search trend data series; second, the new method should be able to deal with a large number of search data series. As a result, a generalized dynamic factor model (GDFM) is adopted to incorporate many keyword variables. An advantage of GDFM is its ability to process high-dimensional data and to use a composite index (Amstad & Potter, 2009). GDFM is commonly adopted in the analysis of economic or financial cycles, but it is seldom used in tourism forecasting. We applied our proposed methodology to predict tourist volumes in Beijing, one of the most renowned travel destinations in the world. By collecting specific search trend data from Baidu including tourism-related keywords ("dining," "lodging," "trip," "traveling," "shopping," and "recreation"), this study empirically tested the method in the forecasting of weekly Beijing tourist volumes from January 2011 to August 2015. The empirical results demonstrate that our method is superior to the benchmark models of an autoregressive model and a model with PCA as a predictor. This study contributes to existing literature in two aspects: first, it validates the performance of the aggregated index from large search trend datasets; second, empirical results demonstrate that the GDFM model is suitable for accurate tourism demand forecasting.

This paper proceeds as follows: Section II briefly reviews the relevant literature. Section III proposes a framework of integrated index construction. Section IV presents our empirical study and research findings. Finally, Section V concludes by discussing the study's contributions and implications for future research.

2. Literature review

In this section, we first review the current studies on tourism demand forecasting. Second, we focus on big data forecasting with search trend data, including the major techniques in keyword and variable selection. We also introduce the generalized dynamic factor models along with their applications. Third, we address the research gap at the end of this section.

2.1. Tourism demand forecasting: data and techniques

Tourism demand forecasting is a well-established research area, and it has been the subject of many studies in the tourism and hospitality field. Song and Li (2008) conducted a detailed literature review on tourist demand forecasting methods and techniques in recent decades. The commonly adopted forecasting techniques are time series, econometric models, artificial intelligence approaches, and hybrid methods.

Time series models predict tourist arrivals based on historical patterns. Many studies used time series models to analyze and forecast tourism demand (Akın, 2015; Athanasopoulos & Hyndman, 2008; Chu, 2008, 2009; Guizzardi & Stacchini, 2015; Gunter & Önder, 2015). The most popular models are autoregressive moving average models (Song & Li, 2008). Econometric models explore the causal relationship between tourist arrivals and influencing factors, which are especially useful when a correlational relationship exists (Song, Romilly, & Liu, 2000; Song & Witt, 2000, 2006; Song, Witt, & Jensen, 2003; Wong, Song, & Chon, 2006; Wong, Song, Witt, & Wu, 2007). Artificial intelligence methods adopt neural networks and support vector machines to model the nonlinear data series (Hadavandi, Ghanbarib, Shahanaghic, & Abbasian-Naghneh, 2011; Pai & Hong, 2005; Pai, Hong, Chang, & Chen, 2006; Palmer, Montano, & Sese, 2006; Palmer et al., 2006). Some studies have proposed a hybrid forecasting by combining econometric and data mining techniques (Pai, Huang, & Lin, 2014; Sun, Wang, Zhang, & Gao, 2016). Researchers also used methods such as meta-analysis and singular spectrum analysis in the modeling and forecasting of tourist arrivals (Hassani, Webstera, Silvaa, & Heravic, 2015; Peng, Song, & Crouch, 2014).

In terms of forecasting accuracy, different models have their own advantages and disadvantages. No single model can consistently outperform others in all situations (Song & Li, 2008). Artificial intelligence techniques can model limited observations. For example, Wang (2004) used fuzzy time series and grey models to predict tourism demand with only 12 data points. Econometric models need a large amount of observations to achieve a higher forecasting accuracy; in comparison, artificial intelligence models lack theoretical foundation for modeling tourism demand (Song & Li, 2008). Researchers are unable to illustrate the detailed influences of each explanative variable on an explained variable. By contrast, econometric models have sound theoretical underpinning, and they can validate the relationship between explained and explanative variables from an economic perspective.

2.2. Big data analytics in tourism research

Big data analytics has become increasingly important in both the academic and the business communities over the past two decades (Chen, Chiang, & Storey, 2012; Xiang, Woeber, & Fesenmaier, 2008). Travelers' decision-making is intrinsically complicated and multidimensional, such as selecting destinations, reserving hotels, planning itineraries, and other activities. The new data sources generated by users based on Internet technology (search engines or social media platforms) have become popular in studying travelers' decision-making and behavior.

Some extant literature has attempted to introduce user-generated content and big data analytics in tourism-related research. With big data sources, tourist arrivals or hotel sales can be forecasted more accurately (Blal & Sturman, 2014). Choi and Varian (2012) investigated the predictive ability of search engine data in travel destination planning. By using keyword search volume data from Google, they increased the prediction accuracy for Hong Kong tourist arrivals from several countries, such as the United States, Canada, Great Britain, and Germany. Yang et al. (2014) predicted hotel demand by combining traditional econometric models with web traffic volumes and demonstrated the use of web volumes in predicting hotel occupancy in a tourist destination.

In addition, these search engine and social media data sources can also help improve customer service, user experience, and satisfaction (Pan, Litvin, & Goldman, 2006). Ye, Law, and Gu (2009) examined the effects of online consumer-generated reviews on hotel room sales. The data were collected from the largest travel website in China. Their research findings indicated a significant relationship between online reviews and the business performance of hotels. Li, Law, Vu, Rong, and Zhao (2015) used online review data from TripAdvisor to identify emergent hotel features of interests of international travelers. Their findings helped hotel managers to gain insights into travelers' interests and to better understand rapid changes in tourist preferences. Ghose, Ipeirotis, and Li (2012) proposed a hotel demand estimation model by combining U.S. hotel reservation data from Travelocity with various social media sources. They applied big data analytics such as text mining, image classification, and social geo-tagging to generate a new ranking system and to provide customers with best-value hotels. Wohlfarth, Clemencon, Roueff, and Casellato (2011) collected Internet-based data and used the descriptive characteristics of flights and text mining to predict travel price changes. The findings help customers to decide when to purchase the best-value tickets. Xiang, Schwartz, Gerdes, and Uysal (2015) exploited the manner in which big data analytics understands the relationship between hotel guest experiences and satisfaction. Big data and text analytics can discover customers' behavior and experiences.

Unlike traditional data sources in tourism research, big data analytics provides a large amount of data without sampling bias. With these new data sources, the academia and industries can better understand consumer behavior in the tourism and hospitality fields.

2.3. Forecasting with search trend data

Researchers have adopted search volume data to forecast many social and economic activities. The forecasted variables include unemployment (Askitas & Zimmermann, 2009), consumption levels (Carrière-Swallow & Labbé, 2013; Vosen & Schmidt, 2011), consumer prices (Choi & Varian, 2012), housing prices (Wu & Brynjolfsson, 2015), and stock prices (Da, Engelberg, & Gao, 2011).

A more recent trend in tourism demand forecasting is forecasting with search trend data from Google and Baidu. Pan et al. (2012) used five travel-related Google search volume data to predict hotel room demand in an autoregressive moving average with an exogenous model. Bangwayo-Skeete and Skeete (2015) used Google search and a mixed data sampling model to improve the forecasting performance of tourist arrivals. Yang et al. (2015) used Baidu and Google search trend data to predict Chinese tourist flows by using autoregressive moving average models and evaluated the performances of the data of the two search queries.

In general, researchers use three main data modeling techniques for search volume data. When the number of relevant search queries is small, researchers include volume data directly (McLaren, 2011; Vosen & Schmidt, 2011). To accomplish the task, researchers use the autoregressive moving average with exogenous variables to construct the forecasting model. When the number of search query data is large, keeping all the variables in the model poses problems because of potential multicollinearity and overfitting issues in the model estimation (Varian, 2014). Thus, constructing indices from a large amount of search query data is a feasible solution. Researchers can use PCA to construct a search index (Li, Shang, Wang, & Ma, 2015). In addition, researchers can aggregate data using data shift and summation in consideration of the lag orders of different types search query data (Yang et al., 2015). Thus, the index is the linear combination of original query data series.

The latter two methods create a search index to be included in the forecasting model. In general, the PCA index is the first factor extracted from the search query data, and the last method is a linear combination of the partials of the original data. However, both indices may fail to represent the dynamics among all search queries. The PCA index is created with the reduction of the originally high dimensions, thus resulting in information loss in the process. The second method using shift and summation linearly combines the search queries, and the aggregated index cannot address the dynamic correlations among search queries. For example, query data series could be dynamically correlated with one another and can be determined by some potentially common factors such as holidays. Therefore, researchers need a new index construction method to process the multidimensional search queries effectively. The new index should be comprehensive and reflect the most relevant information in search queries. It should also depict the common components of search queries to cover different lead and lag information of the search data series.

2.4. Generalized dynamic factor model

One model seems to be an ideal candidate for modeling multidimensional search data. Forni, Hallin, Lippi, and Reichlin (2000) proposed GDFM by extending the dynamic factor models. Let $\{x_{i,t,}i{=}1,...n,t{=}1,...T\}$ be the set of observed variables. Each variable can be modeled as the sum of its common componentχ_i and an idiosyncratic componentξ_i. The common components are driven by a q-dimensional vector of common factors$f_i = (f_{1t}, f_{2t}, \ldots, f_{qt})^{'}$. The model is noted as follows:

$$x_t = \chi_t + \xi_t = B(L)f_t + \xi_t \tag{2.1}$$

$$\chi_{it} = b_{i1}(L)f_{1t} + b_{i2}(L)f_{2t} + \cdots + b_{iq}(L)f_{qt} \tag{2.2}$$

where L is the lag operator, $B(L) = b_{ij}(L), i = 1, 2\ldots, n, j = 1, 2, \ldots, q$ is the set of time-varying factor loadings, and q indicates the number of commonly dynamic factors. Forni et al. (2000) suggested that q is determined through the variance contribution of each component. If the variance contribution rate of the first *i-1* components diverges and component *i* begins to converge, then q is set to i-1. The figure of variance contribution rate is used to determine the value of q. The estimation details of GDFM are found in Forni et al. (2000) and Forni, Hallin, Lippi, and Reichlin (2005).

One feature of GDFM is the estimation of common factors. For example, we aim to predict tourism demand with hundreds of search trend data. In the GDFM framework, these observed variables can be partly explained by common unobserved factors, which are noted as common components. The common components are useful information to explain tourism demand and can be used to aggregate the relevant index.

The model has two distinct superiorities in analyzing data with a large number of variables. First, the model can dynamically update parameters, and thus, it can deal with typically dynamic questions. Second, GDFM allows for cross-correlation among idiosyncratic components. Specifically, GDFM can generate a coincident index to represent the common states of the observed variables. In the existing literature, GDFM is not only capable of reflecting the business cycle or inflation, but it can also forecast the dependent variable with cycles. Recently, GDFM has been widely used in the modeling and forecasting of business cycles (Christophe, 2006), underlying inflation gauges (Amstad & Potter, 2009), and other economic indicators (Forni, Hallin, Lippi, & Reichlin, 2003).

Although other advanced econometric models such as Factor Augmented Vector Autoregressive (FAVAR) model can also analyze a large number of variables with limited factors, they have different advantages. One distinct characteristic of the FAVAR model is that it can analyze the impulse responses of every variable to others. In this study, depicting how each search trend data influence the others in the FAVAR framework is unnecessary. We especially focus on how all search trend data can be combined to optimally improve the forecasting accuracy of tourism demand. Therefore, the FAVAR model is not adopted in this study.

In the field of tourism research, few existing studies used GDFM to model search trend data to predict tourism demand. The tourism industry is dynamically correlated, and the relevant industrial sectors, such as restaurants, hotels, and travel agencies, are also closely correlated with one another. Therefore, GDFM is an ideal candidate for dealing with multiple tourism-related search trend data series.

3. Methodology

After introducing GDFM, we propose a forecasting framework with search trend data, as shown in Fig. 1. This framework describes a modeling process that starts from selecting user-generated data sources, processing data series, and constructing an index, to building forecasting models and evaluating their performances. We present the following six steps of this modeling framework:

(I) Selection of user-generated data sources. Depending on country and culture, different search sources can be used. For example, several search engines in China such as Baidu and Google can provide search services for millions of users. In this empirical study, Baidu search trend data are selected as the major source because Baidu holds the highest market share among all search engines in China. According to the 35th report issued by the China Internet Network Information Center (CNNIC, 2015), Google is less popular in China and has a penetration rate of 27.4%, which is significantly less than the 92.1% penetration rate of Baidu in 2014.

(II) Selection of keywords. The purpose of this step is to define tourism-related keywords. This step intends to follow the mental process of a Chinese traveler who plans to visit a certain city. The traveler first needs to choose the destination. Then, he/she may want to search details about the food, specialties, traffic, weather, hotels, and insights of the place. Thereafter, we define all these aspects, namely, dining, shopping, recreation, lodging, traffic, and insights, as the major factors that travelers consider in their upcoming trips. The search keywords are selected according to these aspects.

(III) Processing of search data. The search data are standardized to 0–100 to eliminate the influence of scales. The value of 100 indicates the highest search query volume, and 0 represents the lowest. Moreover, if the data present an extremely high or low point, then researchers should check the outliers to determine whether the data are affected by one-time events.

(IV) Construction of an index. Through GDFM, a coincident index is constructed to comprehensively represent a traveler's interests in a tourist city. This index reflects the common components of search trend data with a combination of lead and lag information of all search trend data.

(V) Prediction of tourist volumes. Several econometric models are constructed and evaluated. In this research, as a benchmark model, a simple autoregressive moving average model assumes that the current tourist volumes are influenced by past patterns, and thus, it is also considered as a benchmark model in the existing literature (Song & Li, 2008). Furthermore, we conduct another benchmark econometric model that obtains the index using PCA. If the GDFM performs the best among the three models, then we can argue that our method is superior in forecasting tourist volumes.

(VI) Evaluation of forecasting accuracy. To verify the forecasting accuracy of different econometric models, we adopt three criteria: root mean square error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE). The model with the lowest values on the three criteria can be considered as the best forecasting model.

$$MAE = \sum_{i=1}^{n}\left|y_i - \widehat{y}_i\right| \Big/ n \quad MAPE = \frac{100}{n}\sum_{i=1}^{n}\left|y_i - \widehat{y}_i\right| \Big/ y_i \quad RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \widehat{y}_i\right)^2}.$$

where y_i indicates the actual tourist volumes, and $\widehat{y}_i$ is the predicted values of tourist volumes with econometric model.

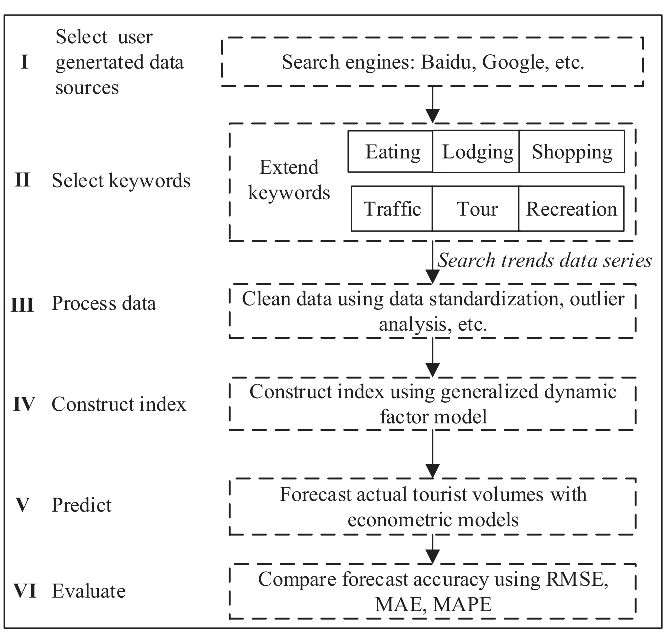

Fig. 1. Forecasting framework with search trend data.

4. Empirical study

4.1. Data

4.1.1. Beijing tourist volumes

Monthly Beijing tourist volumes are collected from the Beijing Tourism Association (BJTA) (2015). The data series, noted as *vis*, ranges from January 2011 to July 2015. Fig. 2 shows the domestic tourist series in Beijing that presents cyclical fluctuations. For convenient modeling, the logs of tourist volumes are computed as *logvis*. However, the search query data are in weekly format. To match up monthly volume data with weekly search query data, the monthly tourist volume series were converted to weekly series by the “quadratic-match average” method (Vogelvang, 2005).

4.1.2. Baidu search trend data

We perform the following steps to select keywords and collect multiple Baidu search trend data to represent travelers' interests and search behavior.

1. Six major aspects of traveling are considered in visiting Beijing, namely, dining, lodging, traffic, tour, shopping, and recreation. These six categories not only reflect the travelers' demand but also represent relevant industries on the supply side.

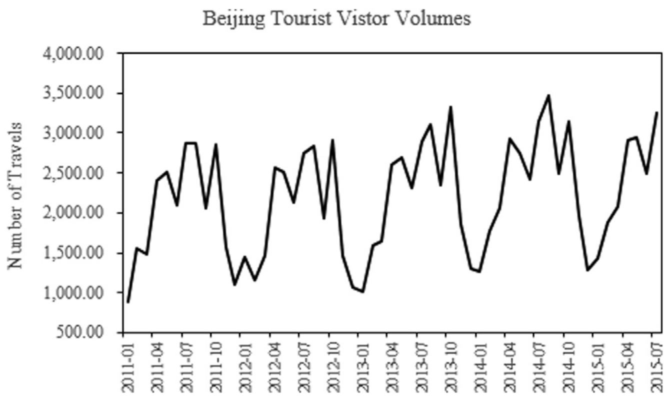

Fig. 2. Monthly Beijing tourist volumes.

2. Several initial keywords in each category are defined using domain knowledge as the seed keywords. These seeds are used in generating more keywords in the following step.
3. The keyword list is extended using the frequently appearing terms on the search engine interface; around 200 keywords are obtained. The purpose of this step is to maximize the possible keyword pool to represent all aspects of tourists' interests in tourism cities.
4. The abovementioned keywords are checked manually on the Baidu search engine to ensure the availability of the search trend data series for download. Baidu does not provide search trend data if the volumes of certain keywords are extremely low. In this process, some keywords with unavailable volume data are eliminated.
5 The volumes for search trend data are obtained. These data series represent weekly frequencies starting from January 1, 2011 to July 25, 2015, as listed in Table 1.

4.1.3. Correlation analysis

Pearson correlations among tourist volumes and all search trend data are computed. Table 1 shows the contemporaneous correlation coefficients between tourist volumes and search trend data. Most search data are positively correlated with tourist volumes. However, some search trend data are poorly correlated (e.g., *Beijing bar* and *Beijing recreation places*). In the existing literature, such as Yang et al. (2015), poorly correlated series below a certain threshold are removed. Unlike their approach, our method keeps the search trend data series in our models to prevent information loss.

4.2. Empirical results

In this section, we construct both GDFM-based and PCA-based indices using their respective models. We then conduct co-integration and Granger causality tests between tourist volumes and the indices. Using these two indices, we construct two forecasting models and evaluate the out-of-sample forecast accuracy of each.

4.2.1. GDFM-based index and PCA-based index

We first construct the GDFM-based index. The figure of variance contribution rate (shown in the appendix) indicates that four factors can explain most of the variance. Accordingly, we set the number of dynamic factors to 4 and the number of lags of the factors to 5 following the Akaike Information Criterion (AIC) (Forni et al., 2000).

Thereafter, we use Eqs. (3.1–3.2) to generate the common components (noted as $C_{i,t}$) in the search query data. The common components are described as follows:

$$C_{i,t} = b_{i1}\left(L^5\right)f_{1t} + b_{i2}\left(L^5\right)f_{2t} + b_{i3}\left(L^5\right)f_{3t} + b_{i4}\left(L^5\right)f_{4t}. \quad (4.1)$$

$$gdfm_t = \sum_{i=1}^{n}\left(C_{i,t}\right)/n. \quad (4.2)$$

Table 1
Search trend data and their correlations with *logvis*.

No.	Search trend data name	Contemporaneous correlation	No.	Search trend data name	Contemporaneous correlation
	Dining			*Lodging*	
1	Beijing food	0.56	8	Beijing hotels	0.64
2	Beijing snack	0.62	9	Beijing accommodation	0.56
3	Beijing specialized food	0.39	10	Beijing restaurant	0.40
4	Beijing specialized snack	0.41	11	Beijing resort	0.50
5	Bejing food guides	0.39	12	Beijing farmhouse	0.42
6	Beijing restaurant	−0.16	13	Hotel booking	0.46
7	Beijing food websites	−0.21	14	Restaurant booking	0.44
			15	Hotel checks	0.11
	Traffic			*Tour*	
16	Beijing flights	0.26	24	Beijing travel	0.63
17	Beijing airports	0.20	25	Beijing travel guides	0.59
18	Beijing shuttle bus	−0.12	26	Beijing travel spots	0.48
19	Shuttle bus schedule	−0.19	27	Beijing fun places	0.58
20	Beijing railway station	0.11	28	Beijing travel agency	0.53
21	Beijing railway tickets booking	−0.17	29	Beijing weather	0.50
22	Beijing railway tickets	−0.25	30	Beijing maps	0.12
23	Bejing bus schedules	−0.12			
	Shopping			*Recreation*	
31	Beijing special products	0.24	40	Beijing recreation	−0.03
32	Beijing shopping	0.28	41	Beijing leisure	−0.09
33	Beijing shopping mall	−0.10	42	Beijing night life	0.22
34	Dashilan street	0.22	43	Beijing show	0.23
35	Panjiayuan center	0.33	44	Beijing bar	0.03
36	Rong bao zhai	0.22	45	Beijing recreation places	−0.03
37	Zhong guan cun	−0.07			
38	Beijing jewelry mall	−0.19			
39	Beijing silk street	0.48			

In the formulas, $C_{i,t}$ is the coincident index (Forni et al., 2000). To construct the GDFM-based index, we require a new index that can represent all the common components in the search query data. Unlike the two methods reviewed in Section 2.3, this new index has two advantages. The GDFM-based index keeps all the original information of search queries. It also considers the lead and lag structures of the common components of search queries to depict their dynamic correlations, as shown in Eq. (4.1). The estimation procedure is conducted in Matlab.

For the PCA-based index, we require factors from the search query data. Let $\{x_{i,t}, i = 1, 2, \ldots, n, t = 1, 2, \cdots, T\}$ be the set of search query data. The extracted factors are noted as follows:

$$pca_t = a_{1i}x_{1t} + a_{2i}x_{2t} + \cdots + a_{ni}x_{nt}. \tag{4.3}$$

The equation suggests that the new factor is the linear combination of the original data. p is the number of factors, and it should be lower than the sample size n. As shown in Table A1, the first factor explains roughly 30% of the variability, which is the largest proportion of the total factors. Accordingly, we construct the PCA-based index by using the first factor.

4.2.2. Co-integration and granger causality tests

Table 2 shows the stability and the Johansen system co-integration tests among the GDFM-based index, PCA-based index, and tourist volumes. These three data series are stable when validated with the augmented Dickey–Fuller test. The co-integration results indicate that the GDFM-based index and Beijing tourist volumes are co-integrated. Similarly, the PCA-based index and Beijing tourist volumes are also co-integrated. Furthermore, a long-term relationship exists between the search trend data and tourist volumes. Therefore, the findings suggest the feasibility of adopting search trend data series in the econometric models.

The purpose of the Granger causality tests is to verify whether the search trend index is predictive of tourist volumes. As shown in Table 3, the GDFM-based index and the PCA-based index are Granger causal of tourist volumes. This finding indicates that search trend data lead the actual Beijing tourist volumes.

4.2.3. Econometric modeling

Three econometric models are constructed to examine the predictive power of the search trend data index. In accordance with the existing literature, Model 1, a simple autoregressive moving averaging model, is selected as benchmark model 1. This model assumes that the current values of time series are affected by its past patterns. Model 2 is constructed as the other benchmark model that incorporates the PCA-based index as the independent variable.

Model 3 uses the GDFM-based index to predict Beijing tourist volumes. Both lags of the GDFM- and PCA-based indices are set to 5 according to the AIC. We equalize the number of independent variables in Models 2 and 3 to evaluate them on the same condition.

Model 1: $Logvis_t = \alpha_1 + \alpha_2 Logvis_{t-1} + \alpha_3 \varepsilon_{t-1} + \varepsilon_t$.

This model serves as benchmark model 1 and considers the first lag of the dependent variable.

Model 2: $Logvis_t = \sum_{i=1}^{5}\beta_{i,t}pca_{t-i} + \beta_6 + \beta_7 Logvis_{t-1} + \beta_8\varepsilon_{t-1} + \varepsilon_t$.

In Model 2, the independent variables in Model 1 are retained, and the lags of pca_t, which is computed using Eq. (4.3), are added. To test the optimal number of principle components, a number of principle components are tested in the forecasting models. However, the forecasting accuracy of the models with more than two components is lower than the model with only the first component. Therefore, only the first principle component is incorporated in Model 2.

Model 3: $Logvis_t = \sum_{i=1}^{5}\gamma_{i,t}gdfm_{t-i} + \gamma_6 + \gamma_7 Logvis_{t-1} + \gamma_8\varepsilon_{t-1} + \varepsilon_t$.

Model 3 incorporates the GDFM-based index obtained from Eq. (4.2). In accordance with Model 2, the lag of this index is set to 5.

In the abovementioned models, $Logvis_t$ indicates the tourist arrivals; $\alpha_1, \alpha_2, \alpha_3$ indicate the coefficients of Model 1; and $\beta_1, \beta_2, \ldots, \beta_8$ and $\gamma_1, \gamma_2, \ldots, \gamma_8$ represent the estimated coefficients of Models 2 and 3, respectively.

Table 4 presents the estimated coefficients and the key measurements of these three models. Adjusted R^2 describes the fitness of the econometric models, and it is a more valid measurement of fitness than R^2. AIC and Schwarz Criterion (SC) are information criteria, which also characterize the performance of the models (Hamilton, 1994). The model with the lowest AIC and SC has the best performance. The bold number in Table 4 indicates the model with the largest goodness-of-fit and the lowest information criteria.

As shown in Table 4, Model 3 with the GDFM-based index has the highest adjusted R^2 and the lowest AIC and SC. Model 3 has an adjusted R^2 that was improved by 1.57% and 0.11% compared with the benchmark and the model with the PCA-based index, respectively.

In terms of the estimated coefficients, all the independent variables are positively correlated with tourist volumes. In Model 3, a 1% increase in the lags of the GDFM-based index leads to an approximately 0.1% increase in the variance of the dependent variable with other variables unchanged. The coefficients of the GDFM-based index in Model 3 are significantly higher than those in Model 2, thereby suggesting that the GDFM-based index possesses greater explanatory power than the PCA-based index.

4.2.4. Out-of-sample forecasting evaluation

Static and dynamic checks on the out-of-sample forecasting power are conducted to evaluate the forecasting accuracy of the models. First, we conduct static one-week and four-week out-of-sample forecasts, in which the sample periods from January 2011 to July 18, 2015 are estimated, and forecast the value on July 25, 2015. The major purpose of the four-week out-of-sample forecasts is to examine whether the model has a relatively long-term forecasting ability.

Second, for the robustness of the evaluation, dynamic rolling

Table 2
Co-integration test results.

Augmented Dickey–Fuller tests				
	t statistics		p value	
logvis	−3.7963		0.0034	
GDFM index	−3.6844		0.0049	
PCA index	−3.3980		0.0120	
Cointegration between *logvis* and *GDFM-based index*				
	Eigenvalue	Trace statistic	Critical value	Prob[b]
None [a]	0.06	28.29	15.49	0.00
At most 1 [a]	0.06	13.42	3.84	0.00
Cointegration between *logvis* and *PCA-based index*				
	Eigenvalue	Trace statistic	Critical value	Prob[b]
None [a]	0.07	18.04	15.49	0.02
At most 1	0.00	0.84	3.84	0.36

[a] Denotes rejection of the hypothesis at the 0.05 level.
[b] MacKinnon–Haug–Michelis (1999) p-values.

Table 3
Granger causality tests between the constructed index and tourist volumes.

Null hypothesis	F-statistic	Prob.
GDFM-based index does not Granger cause *logvis.*	34.08	0.00 [a]
logvis does not Granger cause *GDFM-based index.*	0.40	0.53
PCA-based index does not Granger cause *logvis.*	26.33	0.00 [a]
logvis does not Granger cause *PCA-based index*	0.01	0.94

[a] Indicates the significance level at 1%.

Table 4
Estimation of econometric models.

Model 1		Model 2		Model 3	
Variables	Coefficients	Variables	Coefficients	Variables	Coefficients
		β_1	0.0027 [a]	γ_1	0.0206 [a]
		β_2	0.0024 [a]	γ_2	0.0189 [a]
		β_3	0.0023 [a]	γ_3	0.0184 [a]
		β_4	0.0024 [a]	γ_4	0.0191 [a]
		β_5	0.0019 [a]	γ_5	0.0149 [a]
α_1	7.6943 [a]	β_6	5.9833 [a]	γ_6	5.6913 [a]
α_2	0.8829 [a]	β_7	0.9466 [a]	γ_7	0.9275 [a]
α_3	0.5008 [a]	β_8	0.3591 [a]	γ_8	0.3629 [a]
Adjusted R^2	0.9334	Adjusted R^2	0.9470	Adjusted R^2	**0.9480**
AIC	−1.9348	AIC	−2.2164	AIC	**−2.2356**
SC	−1.8905	SC	−2.0964	SC	**−2.1156**
DW	1.8068	DW	1.9010	DW	1.9027

[a] Indicates the significance level of 1%.

window forecasts are conducted to assess the forecasting accuracy. We set the window length, which is the estimated sample periods, to 180 weeks and run one-week and four-week forecasts. In each forecast, the estimated samples and forecasting points are considered dynamic to test the robustness of the model.

Table 5 shows the RMSE, MAE, and MAPE of these econometric models. To compare the performance of Model 3 with those of the other two models, we compute the improvement rate of the forecasting accuracy from Models 1–2 to Model 3. As shown in Table 5, Model 3 outperforms the benchmark model 1 and Model 2 in all the tests. In the static forecasts, the forecast accuracy of the model with the GDFM-based index improved by an average of at least 50% compared with the model with the PCA-based index. In the dynamic forecasts, the forecast accuracy of Model 3 still improved by 28% and 2.5% compared with the benchmark model 1 and Model 2, respectively. In addition, the short-term forecasting ability is superior to the long-term forecasts. Overall, the GDFM-based index significantly improves the accuracy of tourist volume forecasting.

5. Conclusion and managerial implications

This research proposed a new forecasting framework with search trend data and applied it to the prediction of Beijing tourist volumes. First, we introduce a GDFM that uses the common components of search trend data to construct a more comprehensive index. Second, we compare this new index with a traditional time series model and the PCA-based index model commonly used in past studies. We evaluate the performances of the econometric models with different indices by using static and dynamic tests.

The empirical study indicates that our framework has a more favorable performance than other econometric models. First, a significant co-integration relationship exists between the index and Beijing tourist volumes. Second, Granger causality tests suggest that search trend data lead the actual tourist volumes. Third, we demonstrate that the econometric model with the new index has the best forecasting accuracy in the one-week and four-week forecasts. We also conduct rolling window forecasts to check the robustness. The empirical results validate our framework, which offers a suitable solution for better manipulating large-scale search trend data.

In the context of research using Google or Baidu analytics for tourism forecasting, the primary contribution of this study is to compose the index from large search trend data with a GDFM model. In the study of Yang et al. (2015), the aggregation of an index is computed through the lead–lag orders among these search trend data. When Google or Baidu search trend data possess high dimensions and complicated correlations, our model is demonstrated as an efficient technique.

The proposed framework also has implications for management. Accurate tourist forecasts are crucial for tourism management and planning in the public and business sectors. Search trend data have

Table 5
Forecasting evaluation.

Static out-of-sample forecasting					
Forecasting step = 1 (Forecasting period: July 25, 2015)				Forecasting accuracy improvement rate	
	Model 1	**Model 2**	**Model 3**	**Model 3 vs. Model 1**	**Model 3 vs. Model 2**
RMSE	0.1084	0.0123	**0.0048**	95.57%	60.98%
MAE	0.1084	0.0123	**0.0048**	95.57%	60.98%
MAPE	1.3164	0.1491	**0.0584**	95.56%	60.83%
Forecasting step = 4 Forecasting periods: *July 4–25, 2015*					
	Model 1	Model 2	Model 3	**Model 3 vs. Model 1**	**Model 3 vs. Model 2**
RMSE	0.2659	0.0239	**0.0128**	95.18%	46.42%
MAE	0.2312	0.0180	**0.0123**	94.68%	31.85%
MAPE	2.8387	0.2207	**0.1526**	94.62%	30.82%
Dynamically rolling forecast, window length = 180 weeks					
Forecasting step = 1				**Forecasting accuracy improvement rate**	
	Model 1	**Model 2**	**Model 3**	**Model 3 vs. Model 1**	**Model 3 vs. Model 2**
RMSE	0.0730	0.0574	**0.0566**	22.37%	1.27%
MAE	0.0534	0.0435	**0.0417**	21.91%	4.19%
MAPE	0.6877	0.5645	**0.5414**	21.28%	4.10%
Forecasting step = 4					
	Model 1	Model 2	Model 3	**Model 3 vs. Model 1**	**Model 3 vs. Model 2**
RMSE	0.2409	0.1626	**0.1601**	33.55%	1.57%
MAE	0.2009	0.1289	**0.1281**	36.26%	0.67%
MAPE	2.6068	1.6713	**1.6184**	37.91%	3.16%

The bold numbers indicate the lowest error rates (RMSE, MAE, and MAPE).

been validated to significantly improve the forecasting accuracy of tourism demand, so they can serve as alternative data sources for better tourism management. The GDFM framework is especially useful for modeling high dimensional search engine query data compared to traditional time series models; in this case, high dimensional search trend data reflect many aspects of tourism demand and thus help reduce forecasting errors significantly. The benefit of the reduction of error rates is significant: at least 60% when compared to the benchmark models in the out-of-sample static forecasting; in addition, on the side of cost, search engine query data can be obtained for free from major search engines, such as Google or Baidu. The modeling process takes some time of a researcher but the cost will be reduced if the model can be reused many times for a certain period. Thus, the new model will be likely cost-effective. In practice, managers or analysts can collect various search trend data to compose the index and use the aggregated index to predict the actual tourism demand for strategic decision making. The search index lead time is validated as five weeks ahead of tourist volumes. This demonstrated the travelers' planning behavior to Beijing and also can be used to provide timely forecasting for strategic decision making.

This research has several limitations, and some of them can be investigated for future research. First, we believe the proposed method offers an efficient solution for index aggregation of search trend data, and demonstrated that our model performs best among two benchmark models. However, we cannot conclude that our model is the best among all the competing ones. Various modeling techniques exist, such as artificial intelligence and advanced econometric models, and we did not consider these alterative models. Future research can compare our model with genetic algorithms, artificial neural networks, and vector autoregressive models in tourism demand forecasting. In particular, machine learning approaches can be used to model large and nonlinear datasets, and further investigation needs to be conducted. Second, when the data have different frequency, researchers usually transform weekly data to monthly data at the cost of information loss. In this study, we conduct weekly forecasting models by converting monthly tourist volumes into weekly series to utilize the high-frequency information of search engine data. It is likely that the two data transformation approaches could generate different error rates; however, it still validated the composite search index with better forecasting accuracy. The forecasting evaluation is valid and robust due to the same benchmark and data frequency. Future work can examine the performance of the GDFM index in the monthly forecasting models by transforming the frequency of search engine query data. Furthermore, we expect that the new forecasting framework can be extended to other domains, but we still need further rigorous experiments to examine the scalability of the framework. Future studies can analyze other important indicators, such as tourist volumes in other cities, hotel sales, and flight bookings in the field of tourism and hospitality.

We hope that this study addresses the importance of search index aggregation and will encourage more studies to use Google or Baidu search data for tourism forecasting. Future studies may emphasize interdisciplinary innovation in big data analytics and tourism research.

Acknowledgements

This research was partially supported by grants from the National Natural Science Foundation of China (NSFC No. 71373023 and NSFC No. 41428101) and "New Start" Academic Research Projects of Beijing Union University (Zk10201609).

Appendix A

Table A.1
Results of PCA

Number	Value	Difference	Proportion	Cumulative Value	Cumulative proportion
1	13.7479	1.3265	0.3055	13.7479	0.3055
2	12.4214	8.9336	0.2760	26.1692	0.5815
3	3.4877	0.5389	0.0775	29.6570	0.6590
4	2.9488	1.4649	0.0655	32.6058	0.7246
5	1.4839	0.3688	0.0330	34.0896	0.7575
6	1.1151	0.0711	0.0248	35.2048	0.7823
7	1.0440	0.0675	0.0232	36.2488	0.8055
8	0.976454	0.121597	0.0217	37.22521	0.8272
9	0.854856	0.14662	0.019	38.08007	0.8462
10	0.708236	0.085313	0.0157	38.7883	0.862

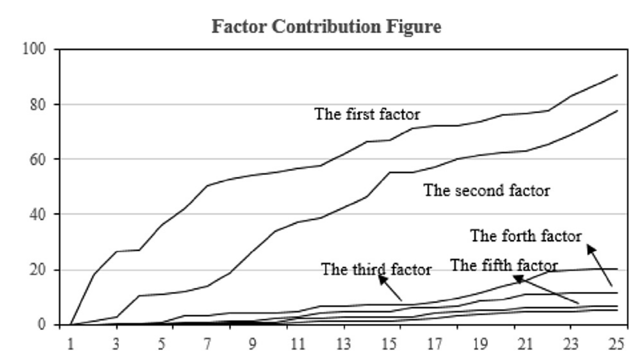

Fig. A.1. Variance contribution rate.

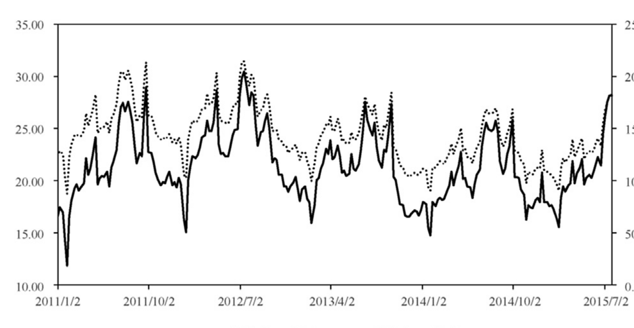

Fig. A.2. GDFM-based index and PCA-based index.

References

Akın, M. (2015). A novel approach to model selection in tourism demand modeling. *Tourism Management, 48*, 64–72.
Amstad, M., & Potter, S. (2009). *Real time underlying inflation gauges for monetary policymakers* (p. 420). FRB of New York Staff Report.
Askitas, N., & Zimmermann, K. F. (2009). Google econometrics and unemployment forecasting. *Applied Economics Quarterly, 55*, 107–120.
Athanasopoulos, G., & Hyndman, R. J. (2008). Modelling and forecasting Australian domestic tourism. *Tourism Management, 29*(1), 19–31.
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.
Beijing Tourism Association (BJTA). (2015). *Beijing statistical information net*. Available online at http://www.bjstats.gov.cn/sjfb/bssj/jdsj/2015.
Blal, I., & Sturman, M. C. (2014). The differential effects of the quality and quantity of online reviews on hotel room sales. *Cornell Hospitality Quarterly, 55*(4), 365–375.
Brynjolfsson, E., Geva, T., & Reichman, S. (2016). Crowd-squared: Amplifying the predictive power of search trend data. *MIS Quarterly, 40*(3).
Carrière-Swallow, Y., & Labbé, F. (2013). Nowcasting with Google trends in an emerging market. *Journal of Forecasting, 32*(4), 289–298.
Chen, H., Chiang Roger, H. L., & Storey, V. C. (2012). Business intelligence and analytics: From big data to big impact. *MIS Quarterly, 36*(4), 1165–1188.
China Internet Network Information Center (CNNIC). (2015). *The penetrate rates of search engines in China of 2014. The 35th statistical report on internet development in China* (pp. 45–46). Available online at https://www.cnnic.cn/hlwfzyj/hlwxzbg/201502/P020150203551802054676.pdf.
Choi, H., & Varian, H. (2012). Predicting the present with google trends. *Economic Record, 88*(s1), 2–9.
Christophe, V. N. (2006). *A generalized dynamic factor model for the Belgian economy: Identification of the business cycle and GDP growth forecasts* (National Bank of Belgium Working Papers).
Chu, F. L. (2008). Analyzing and forecasting tourism demand with ARAR algorithm. *Tourism Management, 29*(6), 1185–1196.
Chu, F. L. (2009). Forecasting tourism demand with ARMA-based methods. *Tourism Management, 30*(5), 740–751.
Da, Z., Engelberg, J., & Gao, P. (2011). In search of attention. *The Journal of Finance, 66*(5), 1461–1499.
Fesenmaier, D. R., Xiang, Z., Pan, B., & Law, R. (2011). A framework of search engine use for travel planning. *Journal of Travel Research, 50*(6), 587–601.
Forni, M., Hallin, M., Lippi, M., & Reichlin, L. (2000). The generalized dynamic-factor model: Identification and estimation. *Review of Economics and Statistics, 82*(4), 540–554.
Forni, M., Hallin, M., Lippi, M., & Reichlin, L. (2003). Do financial variables help forecasting inflation and real activity in the euro area? *Journal of Monetary Economics, 50*(6), 1243–1255.
Forni, M., Hallin, M., Lippi, M., & Reichlin, L. (2005). The generalized dynamic factor model. *Journal of the American Statistical Association, 100*(471), 830–840.
Ghose, A., Ipeirotis, P. G., & Li, B. (2012). Designing ranking systems for hotels on travel search engines by mining user-generated and crowdsourced content. *Marketing Science, 31*(3), 493–520.
Guizzardi, A., & Stacchini, A. (2015). Real-time forecasting regional tourism with business sentiment surveys. *Tourism Management, 47*, 213–223.
Gunter, U., & Önder, I. (2015). Forecasting international city tourism demand for Paris: Accuracy of uni- and multivariate models employing monthly data. *Tourism Management, 46*, 123–135.
Hadavandi, E., Ghanbarib, A., Shahanaghic, K., & Abbasian-Naghneh, S. (2011). Tourist arrival forecasting by evolutionary fuzzy systems. *Tourism Management, 32*(5), 1196–1203.
Hamilton, J. D. (1994). *Time series analysis* (1st ed.). Princeton: Princeton University Press.
Hassani, H., Webstera, A., Silvaa, E. S., & Heravic, S. (2015). Forecasting U.S. tourist arrivals using optimal singular spectrum analysis. *Tourism Management, 46*, 322–335.
Li, G., Law, R., Vu, H. Q., Rong, J., & Zhao, X. R. (2015). Identifying emerging hotel preferences using emerging pattern mining technique. *Tourism Management, 46*, 311–321.
Li, X., Shang, W., Wang, S., & Ma, J. (2015). A MIDAS modelling framework for Chinese inflation index forecast incorporating google search data. *Electronic Commerce Research and Applications, 14*(2), 112–125.
Mayer-Schonberger, V., & Cukier, K. (2013). *Big data: A revolution that will transform how we live, work, and think*. Eamon Dolan/Houghton Mifflin Harcourt.
McLaren, N. (2011). Using internet search data as economic indicators. *Bank of England Quarterly Bulletin, 51*(2).
Pai, P. F., & Hong, W. C. (2005). An improved neural network model in forecasting arrivals. *Annals of Tourism Research, 32*, 1138–1141.
Pai, P. F., Hong, W. C., Chang, P. T., & Chen, C. T. (2006). The application of support vector machines to forecast tourist arrivals in Barbados: An empirical study. *International Journal of Management, 23*, 375–385.
Pai, P. F., Huang, K. C., & Lin, K. P. (2014). Tourism demand forecasting using novel hybrid system. *Expert Systems with Applications, 41*(8), 3691–3702.
Palmer, A., Montano, J. J., & Sese, A. (2006). Designing an artificial neural network for forecasting tourism time series. *Tourism Management, 27*(5), 781–790.
Pan, B., Litvin, S. W., & Goldman, H. (2006). Real users, real trips, and real queries: An analysis of destination search on a search engine. In *Annual conference of travel and tourism research association, Dublin, Ireland*.
Pan, B., Wu, D. C., & Song, H. (2012). Forecasting hotel room demand using search engine data. *Journal of Hospitality and Tourism Technology, 3*(3), 196–210.
Peng, B., Song, H., & Crouch, G. I. (2014). A meta-analysis of international tourism demand forecasting and implications for practice. *Tourism Management, 45*, 181–193.
Song, H., & Li, G. (2008). Tourism demand modelling and forecasting-A review of recent research. *Tourism Management, 29*(2), 203–220.
Song, H., Romilly, P., & Liu, X. (2000). An empirical study of outbound tourism demand in the UK. *Applied Economics, 32*, 611–624.
Song, H., & Witt, S. (2000). *Tourism demand modeling and forecasting: Modern econometric approaches*. Oxford: Pergamon Press.
Song, H., & Witt, S. F. (2006). Forecasting international tourist flows to Macau. *Tourism Management, 27*(2), 214–224.
Song, H., Witt, S. F., & Jensen, T. C. (2003). Tourism forecasting: Accuracy of alternative econometric models. *International Journal of Forecasting, 19*, 123–141.
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.
Varian, H. R. (2014). Big data: New tricks for econometrics. *Journal of Economic Perspectives, 28*(2), 3–27.
Vogelvang, B. (2005). *Econometrics: Theory and applications with eviews*. Pearson Education.
Vosen, S., & Schmidt, T. (2011). Forecasting private consumption: Survey-based indicators vs. Google trends. *Journal of Forecasting, 30*(6), 565–578.
Wang, C. H. (2004). Predicting tourism demand using fuzzy time series and hybrid grey theory. *Tourism Management, 25*, 367–374.
Wohlfarth, T., Clemencon, S., Roueff, F., & Casellato, X. (2011). A data-mining approach to travel price forecasting. In *The 10th international conference on machine learning and applications and workshops, 1* pp. 84–89).
Wong, K. K. F., Song, H., & Chon, K. S. (2006). Bayesian models for tourism demand forecasting. *Tourism Management, 27*(5), 773–780.
Wong, K. K. F., Song, H., Witt, S. F., & Wu, D. C. (2007). Tourism forecasting: To combine or not to combine? *Tourism Management, 28*(4), 1068–1078.
Wu, L., & Brynjolfsson, E. (2015). The future of prediction: How Google searches foreshadow housing prices and sales. In A. Goldfarb, S. M. Greenstein, & C. E. Tucker (Eds.), *Economic analysis of the digital economy* (1st ed., pp. 89–118). Chicago: University of Chicago Press.
Xiang, Z., Schwartz, Z., Gerdes, J. H., Jr., & Uysal, M. (2015). What can big data and text analytics tell us about hotel guest experience and satisfaction? *International Journal of Hospitality Management, 44*, 120–130.
Xiang, Z., Woeber, K., & Fesenmaier, D. (2008). Representation of the online tourism domain in search engines. *Journal of Travel Research, 47*(2), 137–150.
Yang, X., Pan, B., Evans, J. A., & Lv, B. (2015). Forecasting Chinese tourist volume with search engine data. *Tourism Management, 46*, 386–397.
Yang, Y., Pan, B., & Song, H. (2014). Predicting hotel demand using destination marketing organizations' web traffic data. *Journal of Travel Research, 53*(4), 433–447.
Ye, Q., Law, R., & Gu, B. (2009). The impact of online user reviews on hotel room sales. *International Journal of Hospitality Management, 28*(1), 180–182.

Xin Li, Ph.D., is a researcher at e-Tourism Research Center, Institute of Tourism, Beijing Union University. Dr. Li's research interests are big data analytics, econometric modeling, data mining and forecasting. Dr. Li focuses on understanding tourism activities by combing user-generated contents with econometric and machine learning techniques. She also participated in many research projects on monitoring, forecasting, and early-warning of economy and industries in China.

Bing Pan, Ph.D., is an Associate Professor in the Department of Recreation, Park, and Tourism, College of Health and Human Development, Penn State University. His research interests include information technologies in tourism, destination marketing, and search engine marketing.

Rob Law, Ph.D., is a Professor at the School of Hotel and Tourism Management, the Hong Kong Polytechnic University. His research interests are information management, modeling and forecasting, artificial intelligence and technology applications. He has received many research related awards and honors, as well as millions of USD external and internal research grants. Prof. Law serves in different roles for 140 + research journals, and is a chair/committee member of more than 130 international conferences.

Xiankai Huang, Ph.D., is a Professor at Beijing Open University. Prof. Huang's research interests are statistical modeling and tourism forecasting. Prof. Huang has hosted many national programs such as National Key Technology Research and Development Program of the Ministry of Science and Technology of China, with over millions of research grants.