Check for updates

Empirical Research Article

Journal of Travel Research
2020, Vol. 59(1) 52–68
© The Author(s) 2019
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/0047287518824158
journals.sagepub.com/home/jtr
SAGE

Forecasting Tourism Demand with Decomposed Search Cycles

Xin Li, PhD[1] and Rob Law, PhD[2]

Abstract

This study aims to examine whether decomposed search engine data can be used to improve the forecasting accuracy of tourism demand. The methodology was applied to predict monthly tourist arrivals from nine countries to Hong Kong. Search engine data from Google Trends were first decomposed into different components using an ensemble empirical mode decomposition method and then the cyclical components were examined through statistical analysis. Forecasting models with rolling window estimation were implemented to predict the tourist arrivals to Hong Kong. Results indicate the proposed methodology can outperform the benchmark model in the out-of-sample forecasting evaluation of Choi and Varian (2012). The findings also demonstrate that our proposed methodology is superior in forecasting turning points. This study proposes a unique decomposition-based perspective on tourism forecasting using online search engine data.

Keywords

Google search data, ensemble empirical mode decomposition, tourism demand, tourism forecasting

Introduction

The accurate and timely forecasting of tourism demand is imperative for both academia and tourism industries (Akın 2015; Assaf et al. 2019; Chen et al. 2019; Song and Li 2008; Song, Witt, and Li 2008). Researchers seek to improve forecasting accuracy and reduce the errors by utilizing various complicated techniques and appropriate data sets. One major objective of studies on tourism forecasting is to evaluate whether a newly established methodology enhances tourism forecasting compared with the benchmark model (Dergiades, Mavragani, and Pan 2018). Numerous techniques have been proposed to improve the forecasting accuracy of tourism demand, such as time series and econometric models, artificial intelligence algorithms, and hybrid approaches (Chen et al. 2019; Goh and Law 2002; Wang 2004; G. Yu and Schwartz 2006). The accuracy of tourism forecasting is also critical to support business decisions in destination management. In recent years, online data, such as search engine data, web traffic, social media, and mobile data, have gained popularity in forecasting tourism demand (Vu et al. 2018). Online data have unique advantages over traditional data (e.g., survey-based data), including higher frequency (i.e., daily, weekly, and monthly) and capability to represent the behavior of people from various perspectives (Pan and Yang 2017). Consequently, researchers have utilized online data to supplement traditional data and incorporated them into forecasting models. The majority of existing studies demonstrate that online search engine data can improve the forecasting performance of tourism demand (Li et al. 2017; Pan, Wu, and Song 2012; Yang et al. 2015).

From a search behavior perspective, search engine data represent the attention of people to specific topics, such as decisions on travel destinations. People tend to search first for travel destinations, hotels, airlines, and travel agencies before taking their trips (Fesenmaier et al. 2011). The search behavior of people can reflect the future tourism demand to an extent, and thus, search engine data can be viewed as an explanatory variable for tourism demand. Existing studies primarily present empirical evidence to support the hypothesis that search data can improve the forecasting accuracy (Li et al. 2017; Pan and Yang 2017; Yang et al. 2015). However, the reason that why search data has a significant influence on forecasting accuracy remains unclear. Here, we attempt to determine what comprises search data. In particular, we examine the different components of search engine data with a decomposition method. The concept of data decomposition offers a feasible solution to this question. By decomposing the original search engine data series into distinct components, what contributes most to the improvement of the forecasting accuracy can be identified.

[1]Donlinks School of Economics and Management, University of Science and Technology Beijing, Beijing, China
[2]School of Hotel and Tourism Management, The Hong Kong Polytechnic University, Kowloon, Hong Kong

Corresponding Author:
Xin Li, Collaborative Innovation Center of eTourism, Institute of Tourism, Beijing Union University, Beijing, 100101, China.
Email: leexin111@163.com

In recent years, several nonparametric decomposition methods have attracted increasing attention from researchers. Hassani et al. (2015) utilized a singular spectrum analysis that can be regarded as a data decomposition approach to forecast tourist arrivals to the United States. Chen, Lai, and Yeh (2012) adopted a method using empirical model decomposition and a neural network to forecast tourism demand, and results supported the superiority of the decomposition method. The individual components are particularly useful to interpret the original data series better. Zhang et al. (2017) demonstrated that an ensemble empirical model decomposition (EEMD) approach can improve the forecasting accuracy of daily hotel occupancy. Although these studies adopted the concept of decomposition, none applied it to search engine data and analyzed the decomposed components thereof. By decomposing search engine data, we can not only understand the commonly used search engine data but also determine whether the decomposed search data can improve the forecasting accuracy.

Therefore, this study innovatively investigated whether the decomposition of search engine data could improve tourism forecasting. We collected Google search engine data and then utilized an EEMD method to decompose the Google data into different components (named “intrinsic mode functions”). Several statistical analyses were conducted to interpret the components of Google search data. Consequently, the forecasting model with the major components was constructed to examine whether the decomposition could improve the forecasting. To make the empirical results more credible, the current study adopted the same data set used by Choi and Varian (2012) to evaluate the ability of Google data to forecast Hong Kong tourist arrivals from nine countries. The use of the Google data set in the current study has several advantages. First, the results from the same data set can provide a credible benchmark. The data set is open and authentic, so researchers can examine whether the proposed methodology outperforms the benchmark model by using the same data set. Second, by using nine tourist arrivals to Hong Kong, the study can produce robust forecasting results by using rolling window analysis. Finally, the empirical study extended the forecasting by providing a detailed out-of-sample forecasting evaluation, whereas Choi and Varian (2012) only gave the in-sample fit. Therefore, we assert that the proposed methodology emphasizes the efficiency of the decomposed search engine data in out-of-sample forecasting.

The rest of the article is organized as follows. The next section reviews existing studies on forecasting using search engine data and decomposition methods. The Methodology section illustrates the primary methods and framework used in this study, and the subsequent section presents the data and results. The final section presents the conclusions, implications, limitations, and possible future research directions.

Literature Review

Researchers have explored numerous methods to improve the forecasting accuracy of tourism demand, including time series and econometric models, artificial intelligence methods, and forecasting hybrid approaches (Assaf et al. 2019; Chen et al. 2019; Fu et al. 2019; Li, Song, and Witt 2005; Long, Liu, and Song 2019; Witt and Witt 1995; Li et al. 2006; Song and Li 2008; Song, Witt, and Li 2008). With the development of Internet technology and big data analytics, Internet-generated data, such as search engine data and social media data, have attracted attention from academia (Chen, Chiang, and Storey 2012; Pan and Yang 2017). In particular, search engine data have gained increasing popularity, with their distinct advantages of real-time, high frequency, and potential to sensitively capture the behavior of consumers (Yang, Pan, and Song 2014). This section briefly reviews the study about tourism forecasting with search engine data and summarizes the advantages of the decomposition-based methods in improving forecasting accuracy.

Forecasting with Search Engine Data

Given their popularity in measuring the attention from tourists or consumers, search engine data have been considered an effective supplement for traditional data used in forecasting. Researchers have incorporated search engine data into forecasting tasks of various activities, such as epidemic outbreaks, economic growth, unemployment rate, automobile or movie sales, and consumer confidence (Askitas and Zimmermann 2009; Choi and Varian 2012; Ginsberg et al. 2009). In tourism and hospitality literature, the number of studies on forecasting tourist arrivals, hotel sales, occupancy rates, and expenditure is increasing (Dergiades, Mavragani, and Pan 2018; Park, Lee, and Song 2016; Yang et al. 2015). We compared previous studies using search engine data from the following perspectives: predicted variables, search engine data provider, search keywords, data preprocessing, and forecasting models. Table 1 summarizes the selected literature that incorporated search engine data to predict tourism demand.

A significant feature of the stream of studies lies in their use of econometric models. As shown in Table 1, econometric models have taken a dominant role in estimating and forecasting tourism demand. From the perspective of econometrical theories, search engine data represent a type of explanatory variable that has been hypothesized to potentially enhance the estimation of tourism demand (Li et al. 2017). The rationale for this hypothesis is twofold. First, a considerable proportion of people search before they travel, and these online data can be used to measure the demand before the actual tourism demand data are released (Choi and Varian 2012; Fesenmaier et al. 2011). Second, search engine data can

Table 1. Summary of Selected Tourism Forecasting Studies Using Search Engine Data.

References	Predicted Variable	Search Data Frequency	Data Provider	Keywords	Search Data Preprocessing	Forecasting Models
Artola, Pinto, and de Pedraza García (2015)	Tourist arrivals to Spain	Weekly	Google Trends	Different terms related to traveling to Spain	Correlation analysis	ARX model
Bangwayo-Skeete and Skeete (2015)	Tourist arrival to the Caribbean	Monthly	Google Trends	Hotels in and flights to destinations	Data normalization	MIDAS models
Camacho and Pacce (2017)	Travelers to Spain	Weekly	Google Trends	13 tourism-related terms	Correlation analysis	Dynamic factor model
Chamberlin (2010)	Visits to UK	Monthly	Google Trends	Top searches in vacation destinations	Correlation analysis	ARX model
Choi and Varian (2012)	Tourist arrivals to Hong Kong	Monthly	Google Trends	Hong Kong categories	Average index in the first two weeks	Seasonal AR model
Dergiades, Mavragani, and Pan (2018)	Tourist arrivals to Cyprus	Monthly	Web search intensity indices	Multilingual markets with different platforms	Corrected aggregated search engine index	Noncausality testing
Gunter and Önder (2016)	Tourist arrivals to Vienna	Monthly	Google Analytics	10 website traffic indicators	Unit root tests	FAVAR model
Jackman and Naitram (2015)	Tourist arrivals in Barbados	Weekly	Google Trends	Terms from two countries	Correlation analysis	Support vector regressions
Li et al. (2017)	Tourism demand in Beijing, China	Weekly	Baidu Index	45 search terms	Index construction	GFDM model
Önder and Gunter (2016)	Tourist arrivals to Vienna	Monthly	Google Trends	Different languages for "Vienna"	Seasonally adjusted	AR model
Park, Lee, and Song (2016)	Japanese tourist inflow to South Korea	Monthly	Google data	Keyword combination	A systematic method	Multiplicative seasonal ARIMA
Pan and Yang (2017)	Hotel occupancy	Weekly	Google Trends	Charleston hotels	Data normalization	ARIMAX and MSDR
Pan, Wu, and Song (2012)	Hotel rooms	Weekly	Google Trends	Five tourist-related terms	Correlation analysis	ARMAX, ADL, TVP, VAR models
Yang et al. (2015)	Tourism demand in Hainan, China	Weekly	Google/ Baidu Index	10 from Google; 25 from Baidu	Integration, Co-integration	ARX models

Note: ADL = autoregressive distributed lag model; ARMAX = autoregressive moving average with explanatory variables; ARIMAX = autoregressive integrated moving average with explanatory variables; ARX = autoregressive model with explanatory variables; FAVAR = factor-augmented vector autoregressive model; GFDM = generalized dynamic factor model; MIDAS = mixed data sampling approach; MSDR = Markov-switching dynamic regression model; TVP = time-varying parameter model; VAR = vector autoregressive model.

be tested by using traditional econometric theories, such as unit root, Granger causality, and cointegration tests (Li et al. 2017; Yang et al. 2015). In particular, an autoregressive model has generally been considered a benchmark, whereas complex models, such as generalized dynamic factor model, Markov-switching model, Bayesian-based models, and time-varying parameter models, have been proposed as competing models (Assaf et al. 2019; Bangwayo-Skeete and Skeete 2015; Choi and Varian 2012; Li et al. 2017; Pan and Yang 2017). The primary task of the selected studies is to examine whether incorporating search engine data can improve forecasting accuracy measured by reducing forecasting errors, such as root mean squared error, mean absolute error, and mean absolute percentage error (Gunter and Önder 2016; Li et al. 2017).

As demonstrated in the existing literature, the most commonly used search engine data are provided by Google and Baidu search engines, particularly, Google Trends and Baidu Index. A distinct difference between Google and Baidu search data is that Google data are normalized index rather than volume data (Choi and Varian 2012; Yang et al. 2015). However, existing studies have not presented evidence that Google index reduces forecasting power. Therefore, the choice of search engines largely depends on the penetration rates that search engines have in a specific country. For example, the majority of users in the Chinese mainland prefer to use the Baidu search engine, whereas most users in Hong Kong rely on the Google search engine. High penetration rates indicate the generated search engine data are representative (Li et al. 2015). Yang et al. (2015) suggested that Baidu search data outperformed Google data because of the popularity of the Baidu search engine in the Chinese mainland. Google search data are generally used in tourism forecasting in various countries because of the popularity of Google as a search engine (Purcell, Brenner, and Rainie 2012). An advantage of Google Trends is that it provides relevant categories and correlation analysis. In the field of tourism, Google offers “travel,” “vacation destination,” “hotels and accommodation,” “air travel,” and “cruises and charters” (Chamberlin 2010). In comparing the performances of Google and Baidu in predicting the Chinese tourism demand, Yang et al. (2015) found the Baidu search data are superior for improving forecasting accuracy. Önder and Gunter (2016) used a search for “Vienna” with different languages to predict tourism demand for Vienna.

Search engine data are collected daily, weekly, and monthly, whereas tourism demand data sets are generally monthly and yearly series. In existing studies, two solutions are offered to handle the inconsistency between data frequencies. The first is to transform weekly into monthly data by taking the average values in an entire week or the first two weeks (Choi and Varian 2012). The second is to directly use the weekly data by adopting a mixed data sampling approach (MIDAS) (Bangwayo-Skeete and Skeete 2015). A MIDAS model can handle data with different frequencies in an econometric model, although it has a relatively complicated estimation process compared with AR models.

Selecting keywords and preprocessing search data are important steps when modeling search engine data (Brynjolfsson, Geva, and Reichman 2016). Researchers have tended to acquire relevant and representative keywords. Li et al. (2015) evaluated two types of method to select keywords. The first one is to directly use Google-related categories provided by Google Trends (Goel et al. 2010; Vosen and Schmidt 2011). Choi and Varian (2012) investigated how Google data can improve the forecasting of tourist arrivals by collecting data from Google categories. Ginsberg et al. (2009) found that 45 search queries are highly suitable for the weekly influenza detection. However, in tourism and hospitality industries, researchers need to balance the number of search engine data and the complexity of the estimated model to achieve a substantial forecasting accuracy. Therefore, researchers have used the second method to select keywords, that is, few single keywords or a composite index is selected from large keyword groups to reduce the dimension of search engine data (Choi and Varian 2012; Li et al. 2017; Park, Lee, and Song 2016; Yang et al. 2015). Pan, Wu, and Song (2012) used five search engine data sets that reflect the travel and lodging needs for Charleston to predict the weekly demand for hotel rooms.

Forecasting with Decomposition Methods

Decomposition methods have been demonstrated to reduce model complexity effectively and improve the forecasting accuracy significantly (Shabri 2016; Wu and Huang 2009; Zhang et al. 2017). A core concept of the decomposition method is to divide and conquer (Huang et al. 1998). For example, one data series can be decomposed into several components by using different decomposition techniques (Frechtling 2012; Gouveia and Rodrigues 2005). Common decomposition techniques include filter, spectral, singular spectrum analysis (SSA), empirical mode decomposition (EMD), and EEMD. From the perspective of econometrics, a data series contains integrated information on trends, cycles, and seasonal patterns (Hamilton 1994). By adopting a decomposition method, we can obtain a much simpler data series that indicates specific components, such as trend and cycle (Frechtling 2012; Gouveia and Rodrigues 2005).

Coshall (2000) applied spectral analysis to specify the cyclical patterns of international tourism flows and provided a new support for detecting lead or lag relationships between time series data. Cho (2001) used a multiplicative decomposition method to deconstruct tourist arrival data into trend, cyclical factor, seasonal factor, and random error. Nguyen et al. (2013) applied Fourier series to modify the residuals of the Grey model to forecast international tourism demand in Taiwan, which suggested that the residual modification can improve forecasting accuracy.

Decomposition-based SSA methods have numerous applications in tourism forecasting given their advantages in complex data analysis. Beneki, Eeckels, and Leon (2012) applied SSA, a nonparametric method to extract different components, to predict UK tourism income. Their study suggested the superior performance of the decomposition method in reducing the forecasting errors compared with traditional econometric models, such as the seasonal ARIMA model and time-varying state-space models. Hassani et al. (2015) used SSA to predict monthly tourism arrivals from 1996 to 2012 and compared the results with those obtained by ARIMA, exponential smoothing, and neural networks.

Huang et al. (1998) proposed EMD as an effective approach for analyzing complicated data and applied the approach to various academic fields, including economy, finance, and social science (Xu, Qi, and Hua 2010; Yu, Wang, and Lai 2008). For tourism and hospitality industries, Chen, Lai, and Yeh (2012) used an empirical mode decomposition method to acquire data components and individually predicted them with neural networks. Their findings suggested that the decomposition method outperforms the single network model and the ARIMA benchmark model. Yahya, Samsudin, and Shabri (2017) utilized a method based on EMD and a neural network approach to predict the monthly tourist arrivals from Singapore and Indonesia from 2000 to 2013 and reported that their proposed method presented a better forecasting accuracy than those of methods without EMD. Zhang et al. (2017) applied an EMD method to predict the daily hotel occupancy, which improved the accuracy of the short-term forecasting.

The review of the literature shows search engine data have been regarded as a new influencing factor in tourism forecasting. Furthermore, decomposition methods have been used to forecast tourist arrival data to reduce the complexity and improve forecasting accuracy. However, none of the existing studies have combined the advantages of both search engine data and decomposition methods. Search engine data are complicated and unique. The data differ from traditional data because they can reflect the emerging Internet-dependent behavior of people in the big data era. However, the components of search engine data are still undetermined. By directly applying the EEMD method to decompose the search engine data, this study can generate different individual components that are useful to identify the patterns of search engine data. Therefore, the primary investigation of this article focuses on whether the decomposed individual components of search engine data can improve the forecasting accuracy of tourism demand.

Methodology

An EEMD-based framework was proposed to analyze the search engine data, and the consecutive steps, which include specific algorithms, were executed. In this section, the EEMD and forecast evaluation methods are briefly reviewed, and the framework integrating the following steps is presented.

Decomposition of the Search Data

We applied an EEMD method to decompose the search data series at this step. EEMD is an effective method that improves the performance of EMD by adding the white noise to assist in signal extraction (Huang et al. 1998; Wu and Huang 2009; G. Zhang et al. 2017). This method aims to obtain the individual intrinsic mode functions (IMFs) and residuals. EEMD was developed as the following steps (Wu and Huang 2009):

1. Add the data series of white noise to the targeted data.

$$x_i(t) = x(t) + w_i(t),$$

where $x(t)$ is the original data and $w_i(t)$ is the added white noise data series.

2. Decompose the data with the white noise into IMFs.

$$x_i(t) = \sum_{j=1}^{n} c_j + r_n,$$

where c_j denotes the components after n number of IMFs are extracted, r_n is the residual after the decomposition process.

3. Repeat the first two steps iteratively but with a different white noise data series each time.
4. Obtain the (ensemble) means of the corresponding IMFs as the final result.

Interpretation of the IMFs

Several IMFs (noted as IMF_1, IMF_2, and IMF_n) and the residual are generated after the decomposition process. The number of the decomposed IMFs are relevant with the length of data. According to the algorithm proposed by of Huang et al. (1998), the number of IMFs is determined by the following formula:

$$N = \log M - 1$$

where N and M indicate the number of IMFs and the length of data.

The IMFs have different frequencies from high to low frequency. IMF_1 has the highest frequency, and IMF_n has the lowest frequency. Each IMF can reflect specific information such as major and minor cycles contained by the original data. Wu and Huang (2009) suggested that the IMFs with different frequencies can clearly catch the essence of the data. In addition, the decomposed residual indicates the trends of the original data.

Preliminary Analysis

To examine how the IMFs were correlated with the original data series, three measures are constructed including the mean period, Pearson product–moment correlation coefficients, and the variance percentage of each component.

The mean period measures how many peaks or valleys the time series encounters on average during the entire period. The matrix can capture the periodic structure of the data series. This measure is computed using the following formula:

$$T = N / N_1,$$

where N is the length of the data series and N_1 indicates the number of extreme maxima or minima during the period.

The Pearson product–moment correlation coefficient reflects how each IMF is connected to the original search data series. The significance is computed at the 5% level. The variance percentage of each component explains the contribution of each IMF to the total volatility of the search engine data and the sum of the components.

Forecasting and Evaluation

After the decomposition of the search data and the preliminary analysis of individual components, the forecasting models were implemented to examine the forecasting performance. The AR model was first constructed as a benchmark. Subsequently, a set of autoregressive models with explanatory variables (ARX) was estimated by incorporating the independent components of the search engine data.

The benchmark model:

$$y_{t=} \propto + \sum_{i=0}^{j} \beta_i y_{t-i} + \varepsilon_t,$$

ARX model:

$$y_{t=} \propto + \sum_{i=0}^{j} \beta_i y_{t-i} + \sum_{i=0}^{j} \gamma_i x_{t-i} + \varepsilon_t,$$

where y is the predicted variable and x is the explanatory variable. ARIMA models can also be established given the different data characteristics. This step is mainly intended to conduct the out-of-sample forecasting, and thus, the estimated in-sample fitness was not the primary focus. The relevant evaluation was conducted in the next step.

Evaluation was used to examine the accuracy of the out-of-sample forecasting. To achieve the robustness in the evaluation results, we adopted the following research design. First, nine different data sets were used in the robustness checks, as illustrated in the next section. We also employed rolling window estimation and forecasting. The forecasting experiments were designed as follows:

Let window length be L and sample length be N. The forecasting models were run $N - L$ times, and $N - L$ number of predicted values were obtained.

Subsequently, mean absolute percentage error (MAPE) and improvement ratio (IR) as measures of forecasting accuracy were computed using the acquired predicted values. The measures are defined as follows:

$$MAPE = {}^{1}\!\Big/ M \sum_{i=1}^{M} \left(\left| \widehat{y_i} - y_i \right| \Big/ y_i \right) \times 100\%,$$

$$IR = \frac{MAPE_{ARX} - MAPE_{Benchmarck}}{MAPE_{Benchmarck}} \times 100\%.$$

IR indicated that the extended models with decomposed search data improved forecasting accuracy compared with the benchmark model.

Accordingly, a framework combining the models was established, as presented in Figure 1. In this framework, the search engine data are decomposed into individual components. Then, the components undergo several econometric tests, such as correlation and variance analyses. Forecasting and evaluation are conducted to compare the predictive ability of the proposed methods. The proposed methodology can answer the following questions: first, what components do the search engine data represent? Second, can an individual component influence the performance of tourism demand forecasting?

Data and Results

To investigate the predictive ability of the search engine data components in improving the forecasting accuracy of tourism demand, we chose tourist arrival data to Hong Kong from nine countries: the United States, Canada, Great Britain, Germany, France, Italy, Australia, Japan, and India. These data were first collected and examined by Choi and Varian (2012). We utilized the data set in the empirical study for three reasons. First, we can compare the results obtained using the proposed methods with those obtained by Choi and Varian (2012); their work provides a sound benchmark. Second, the entire data set can be used for robustness checks by including search engine data and tourist arrivals to Hong Kong from nine countries. The generalizability of the proposed methodology in forecasting of tourist arrivals can be verified by using the same data set. In this section, the relevant data are described, and the decomposition results obtained by using the proposed methodology are presented. The models were used to estimate and forecast the tourist arrivals with the decomposed search cycles. Forecasting accuracy was evaluated by analyzing the specific error reductions.

Data Description

The two types of collected data are described in this subsection. The first type of data includes the actual tourist arrival

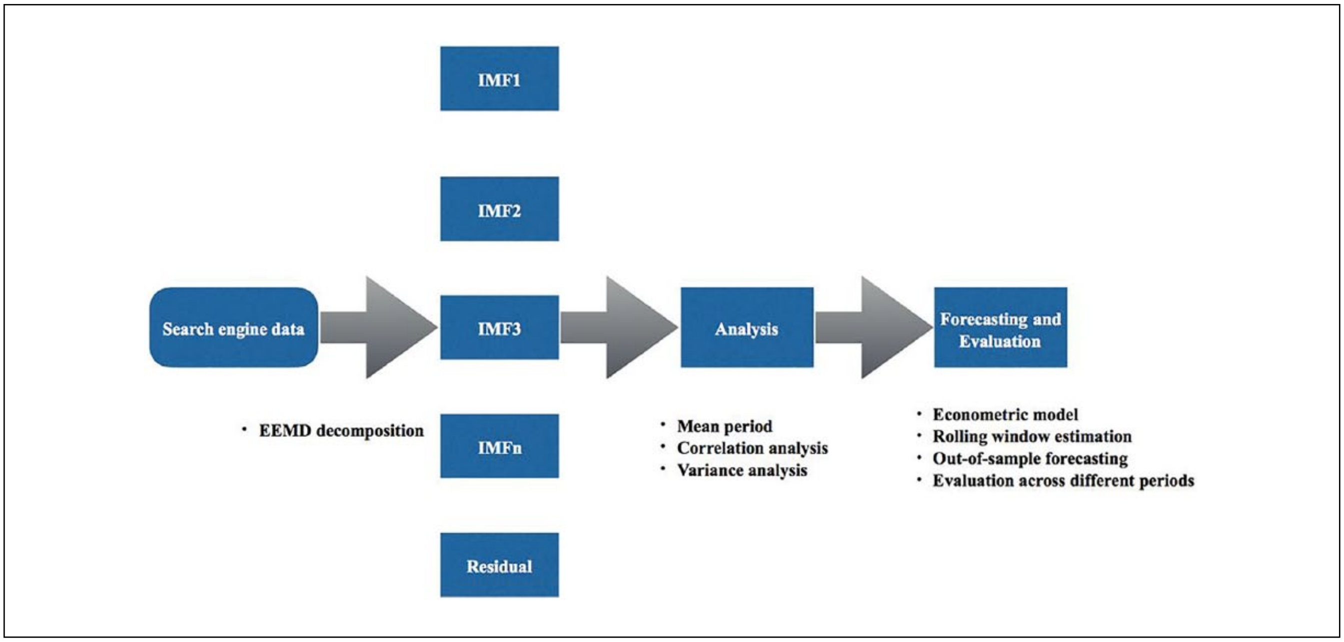

Figure 1. Ensemble empirical decomposition-based methodology.
Note: "US_Search" and "US" indicate Google data from the United States and actual tourist arrival data from United States to Hong Kong, respectively.

data to Hong Kong from nine countries. The data were downloaded from Hong Kong Tourism Board. The second type of data is the Google search query index (Google data), which was acquired from Google Trends. Google Trends allows for the observation of one of the subcategories named "Hong Kong" under Vacation Destinations (Choi and Varian 2012). Accordingly, the search query index for "Hong Kong" can be acquired by country of origin, including the United States, Canada, Germany, and so on. To make the data consistent with those used by Choi and Varian (2012), we selected Google data from nine countries mentioned.

The data series start from January 2004 to February 2011 at the monthly frequency. The actual tourist arrival data are released with a roughly four-week lag. The Google data present the average index in the first two weeks of the month, thereby providing a six-week lead in the forecasting.

The abbreviations US, CA, GB, DE, FR, IT, AU, JP, and IN were used to represent the nine counties. For instance, "US_Search" and "US" indicate the relevant Google search query data from the US and the actual tourist arrivals to Hong Kong from the US, respectively. Figure 2 illustrates the nine sets of data, including tourist arrivals to Hong Kong and the Google search engine data. Table 2 shows the data description.

The descriptive results showed all tourist arrival data presented distinct seasonal patterns, whereas the search engine data showed different complicated characteristics. Google search data were used to predict tourist arrivals to Hong Kong for two reasons. First, existing literature provides evidence that search engine data can measure people's attention and help improve the forecasting accuracy (Engelberg and Gao 2011; Li et al. 2017; Yang et al. 2015). For instance, Google data from the United States to Hong Kong indicate people in the United States show interest or attention on traveling to Hong Kong by using the Google search engine. Although not all searches for Hong Kong can positively influence the actual tourist arrivals, the search data can represent the overall popularity (Choi and Varian 2012). Second, we used a standard Granger causality test, which is commonly used in existing studies (Li et al. 2017; Yang et al. 2015), to examine whether Google data have the predictive ability. The result showed the Google data were the Granger causality of the tourist arrivals data at the 1% level of significance. Accordingly, the nine independent Google search data series were considered explanatory variables to predict monthly tourist arrivals to Hong Kong from nine different countries.

Results of Decomposition

Following the proposed methodology, we applied the EEMD method to decompose the original search data series to individual components. For each search data series, five IMFs and one residual item were generated using the R program. In accordance with the algorithm proposed by Huang et al. (1998), the total number of IMFs was determined by using the formula $\log_2 N - 1$, where N indicates the length of the sample. Therefore, the number of IMFs was restricted to 5. The decomposed data series were IMF1, IMF2, IMF3, IMF4, IMF5, and residual. Figure 3 depicts the components of the Google search engine data of tourist arrivals to Hong Kong from the United States.

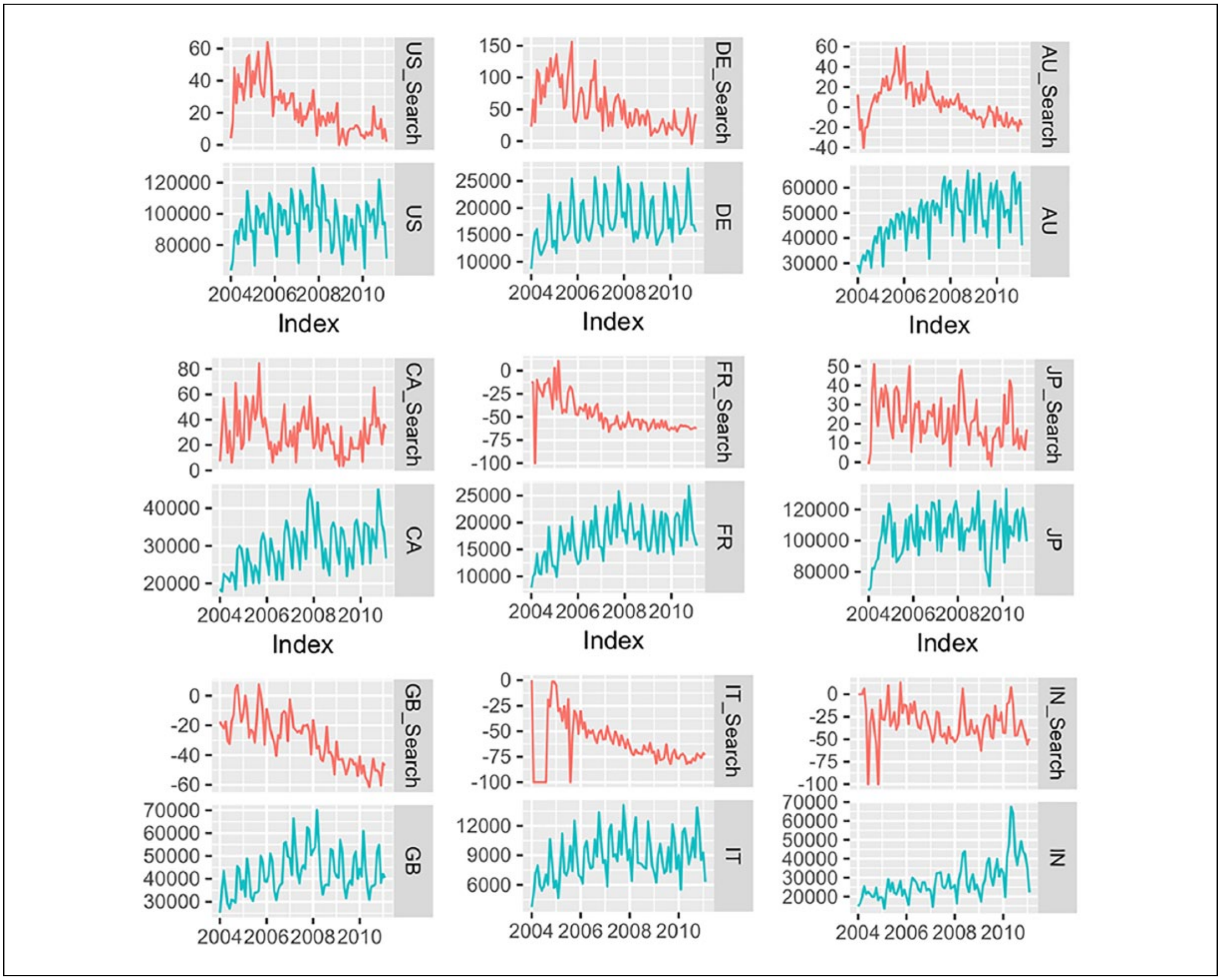

Figure 2. Actual tourist arrivals to Hong Kong and search engine data.

Table 2. Data Description.

	US	US_Search	CA	CA_Search	GB	GB_Search
Mean	94,633	22.35	29,292	29.49	42,670	−29.81
Median	94,150	20	29,251	30.25	40,714	−29.5
Min	63,632	0	17,834	3.5	25,229	−61.5
Max	129,451	64	45,148	84.5	70,102	7.5
	DE	DE_Search	FR	FR_Search	IT	IT_Search
Mean	17,688	53.39	17,270	−47.49	3,736	−61.36
Median	16,476	46	16,980	−53.25	8,447	−65.75
Min	8,716	−4.5	7,848	−100	3,736	−100
Max	27,631	156	26,818	10.5	14,127	0
	AU	AR_Search	JP	JP_Search	IN	IN_Search
Mean	48,422	3.45	105,015	21.32	28,313	−31.07
Median	49,382	1.5	106,618	19.5	25,864	−32.5
Min	26,341	−40.5	68,062	−2	13,593	−100
Max	66,742	60.5	133,252	51	67,439	13

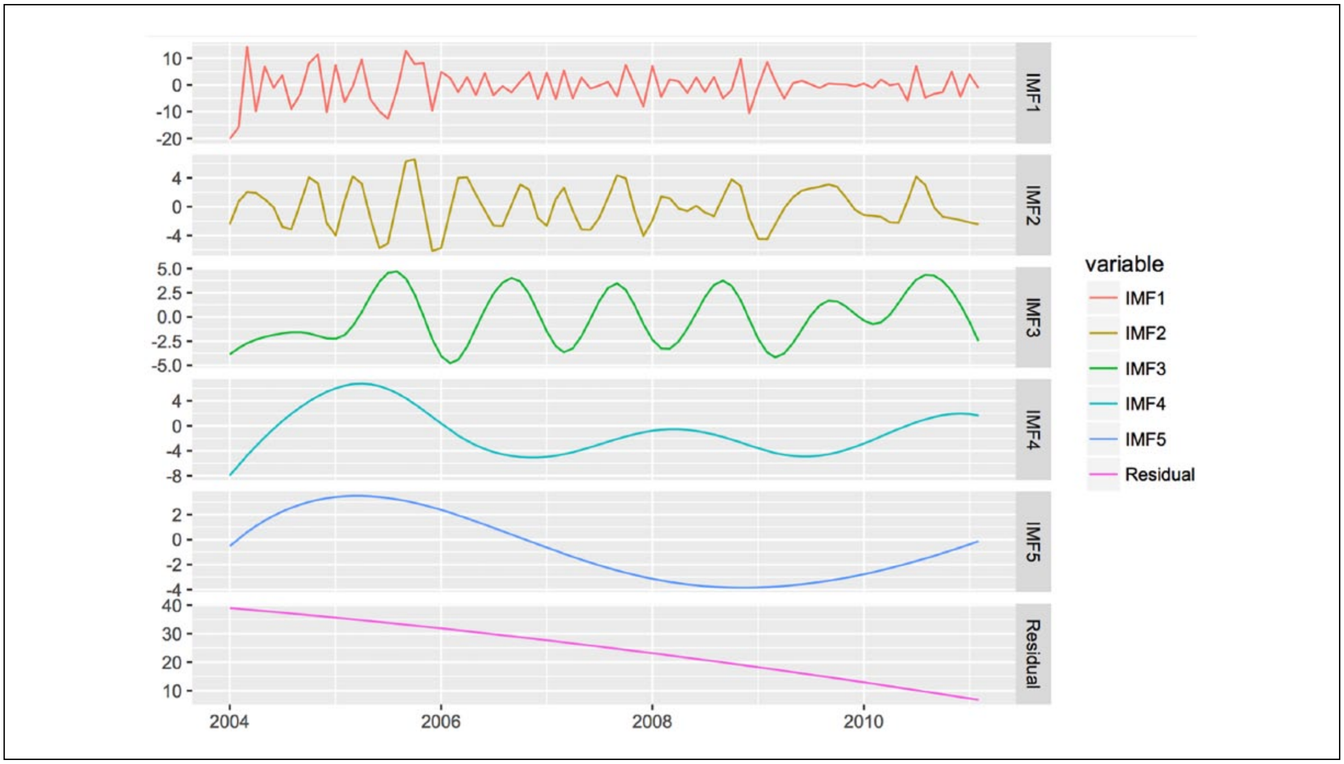

Figure 3. Components of US search data.

After obtaining the decomposed IMFs, we computed the following measures to depict how each IMF is correlated with the original search engine data. These measures include mean period, Pearson correlation, percentage of variance to the sum of the decomposed series, and percentage of variance to the original data series. Table 3 summarizes the measurements of each decomposed IMF.

The results of the decomposition suggested the various representations of each component of the search engine data set. First, from the perspective of the semantic explanation of the components, the major distinction of these components is their frequency. The components of all search engine data sets presented distinctly different frequencies. IMF1 had the highest frequency, whereas IMF5 had the lowest frequency. The higher the frequency of the IMF, the more information and inevitable noise it has. The individual components reflect the trends, cycles, and seasonal patterns of the search engine data. Using the formula for the mean period, we compute that IMF3 showed an average period at approximately 12 months in all search engine data. Therefore, IMF3 can be identified as the major cycle of the search engine data, and it reflects the primary peaks and valleys. Similarly, IMF1 and IMF2 can be considered the minor cycles of the search engine data, with additional distinct peaks and valleys in the period. IMFs 4 and 5 contained the least information and can be regarded as modest fluctuations. By contrast, the residual was a deterministic long-term behavior that can reflect long-term trends in the tourism market. Huang et al. (1998) likewise argued the residual component can be considered a deterministic long-term behavior.

Second, from the perspective of correlation coefficients, nearly all components showed significantly positive relationships to the original search engine data. For instance, IMF1 had an approximately 50% correlation with search engine data from the US. The correlation coefficients represented connections between the original search data and its components.

Third, we computed two types of variance percentage of the components, including the IMFs and residual. The first variance percentage describes how each component contributed to the total volatility of the IMFs and residual. Therefore, the sum of these variances equals 1. The second percentage represents the contribution of an individual component to the original search engine data. The results showed the sum of variances of the IMFs and that the residual was not equal to 1. Zhang, Lai, and Wang (2008) explained the reason for such an outcome lies in the factors, such as the nonlinearity of the original time series.

Accordingly, we obtained the primary distinct cycles (i.e., IMF3) of the search engine data and the cyclical component of the tourist arrival data and compared the two sets of cycles. Figure 4 shows the major cycles of the search engine data and the tourist arrival data.

Table 3. Summary of the Components of the Search Data Sets.

	Mean Period (Month)	Pearson Correlation	Percentage of Variance to the Sum of the Decomposed Series	Percentage of Variance to the Original Search Data
US_Search				
IMF1	2.69	0.4671*	25.62	18.23
IMF2	6.62	0.3669*	4.64	3.30
IMF3	12.29	0.04	3.88	2.76
IMF4	28.67	0.5758*	7.87	5.60
IMF5	43.00	0.7509*	5.18	3.68
Residual		0.7371*	52.81	37.58
CA_Search				
IMF1	2.69	0.6010*	40.61	38.58
IMF2	5.38	0.4191*	18.57	17.65
IMF3	12.29	0.4233*	14.07	13.37
IMF4	28.67	0.4800*	22.14	21.04
IMF5	43.00	0.3306*	2.21	2.10
Residual		0.2073*	2.39	2.27
GB_Search				
IMF1	3.19	0.4559*	17.38	11.98
IMF2	6.62	0.4796*	10.15	7.00
IMF3	12.29	0.2098*	3.12	2.15
IMF4	28.67	0.4467*	3.22	2.22
IMF5	43.00	0.7545*	1.31	0.90
Residual		0.8176*	64.82	44.69
DE_Search				
IMF1	3.44	0.4849*	34.58	27.34
IMF2	7.17	0.3356*	15.07	11.91
IMF3	14.33	0.2535*	9.21	7.28
IMF4	28.67	0.5734*	4.21	3.33
IMF5	43.00	0.6854*	3.85	3.04
Residual		0.6655*	33.08	26.15
FR_Search				
IMF1	2.87	0.5650*	49.39	39.74
IMF2	5.38	0.2756*	6.87	5.53
IMF3	14.33	0.14	3.95	3.18
IMF4	43.00	0.5720*	12.92	10.39
IMF5	43.00	0.6682*	2.01	1.62
Residual		0.6423*	24.86	20.01
IT_Search				
IMF1	2.87	0.3308*	20.99	17.27
IMF2	6.62	0.5446*	28.37	23.35
IMF3	12.29	0.6033*	31.13	25.62
IMF4	28.67	0.3900*	3.67	3.02
IMF5	43.00	0.5012*	3.29	2.71
Residual		0.4410*	12.55	10.33
AU_Search				
IMF1	2.69	0.3082*	19.11	14.38
IMF2	5.73	0.2950*	6.19	4.66
IMF3	14.33	0.5508*	28.44	21.40
IMF4	28.67	0.7466*	6.80	5.12
IMF5	43.00	0.6546*	1.25	0.94
Residual		0.6205*	38.21	28.75
GP_Search				
IMF1	2.97	0.7238*	54.36	48.64

(continued)

Table 3. (continued)

	Mean Period (Month)	Pearson Correlation	Percentage of Variance to the Sum of the Decomposed Series	Percentage of Variance to the Original Search Data
IMF2	6.62	0.5202*	14.89	13.32
IMF3	12.29	0.4149*	13.49	12.07
IMF4	28.67	0.20	2.42	2.17
IMF5	43.00	0.3656*	0.11	0.10
Residual		0.3379*	14.73	13.18
IN_Search				
IMF1	3.58	0.5942*	47.45	53.44
IMF2	6.62	0.5838*	39.25	44.20
IMF3	14.33	0.3595*	7.14	8.04
IMF4	28.67	0.20	4.61	5.19
IMF5	43.00	0.2224*	0.28	0.31
Residual		0.12	1.28	1.44

*The correlation is significant at the 5% level (two-tailed).

As indicated in Figure 4, the tourist arrival data and search engine data presented a significant similarity after decomposition. Therefore, the decomposed IMF3 cycles were used for the subsequent forecasting and evaluation.

Analysis of Forecasting Results

After the decomposition of the search engine data, the decomposed cycles were incorporated into a basic seasonal AR model. The benchmark was the ARX model with the undecomposed search engine data used by Choi and Varian (2012). The comparative EEMD–ARX model was established by adding the decomposed cycles as the exogenous variable. The objective was to investigate whether the new model with the decomposed cycles can improve the forecasting accuracy. The models have the following forms.

Benchmark model:

$$y_t = \alpha y_{t-1} + \beta y_{t-12} + \gamma x_t + e_t.$$

EEMD-ARX model:

$$y_t = \alpha y_{t-1} + \beta y_{t-12} + \gamma x\_c_t + e_t.$$

The difference between the two models is that the benchmark model uses Google data (x_t), whereas the new model extracts the decomposed cycles (x_c_t) as the exogenous variable. The incorporation of the decomposed search data can provide an approximately six-week lead in terms of forecasting.

We estimated the models for each country and conducted the dynamic out-of-sample forecasting to compare the predicted values with their actual values. To examine the predictive ability of the decomposed cycles, we first conducted one-step-ahead out-of-sample forecasting, which predicted tourist arrival data in October 2010 by using the data from January 2004 to September 2010. Table 4 lists the forecasting results.

The results in the forecasting experiments showed the EEMD–ARX models outperformed the benchmark models with different improvement ratios. The average improvement of the proposed methodology for forecasting accuracy was 27.8%. Even though the results suggested the proposed decomposition-based methodology can reduce forecasting errors, we investigated further the reason for the effectiveness of the proposed methodology. Therefore, error reduction analysis was conducted. In the next subsection, detailed examples are provided to compare the actual values, benchmark models, and proposed models.

Evaluation of Error Reduction

We compared the forecasting errors of the proposed model and the benchmark to further validate the robustness of the proposed decomposition-based methodology. We designed a dynamic out-of-sample forecasting with rolling window estimation. To emphasize the short-term forecasting ability of the methodology, we conducted a one-month-ahead forecasting process. In the forecasting procedure, we set the first window length to 20 months and enlarged the length during the rolling window process. For example, we used the data from January 2004 to August 2005 to predict the values in September 2005. We then continued to include the sample in September 2005 to forecast the values in October 2005. The last model used data from January 2004 to January 2011 to predict the tourist arrivals in February 2011. In total, the dynamic models with different sample periods were validated 66 times, indicating the robustness of the forecasting results.

After the forecasting process, we obtained the predicted values from September 2005 to February 2011. Therefore, the average forecasting errors can be computed using the

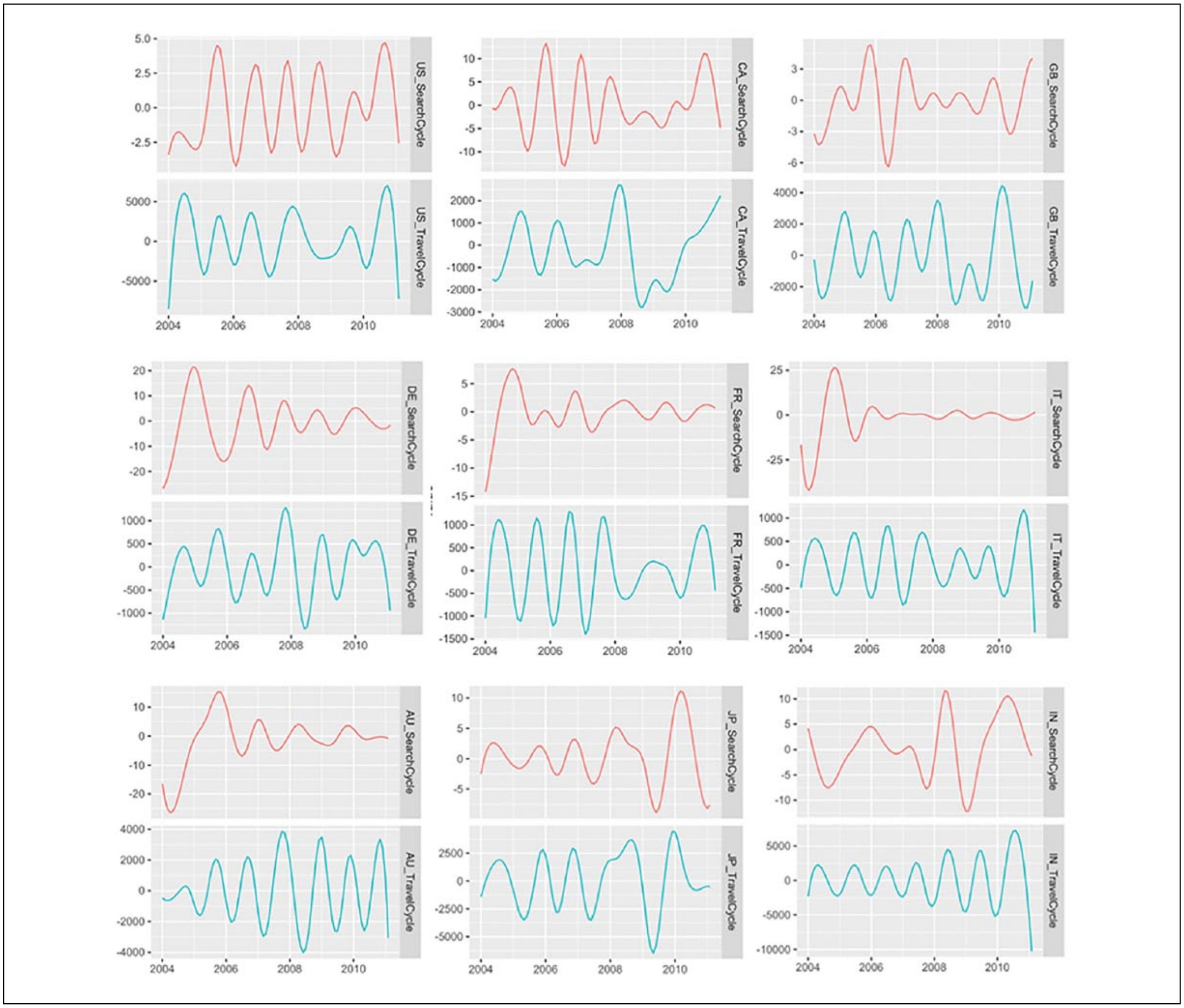

Figure 4. Decomposed cycles of the search engine data and tourist arrival data.

Table 4. One-Step-Ahead Out-of-Sample Forecasting Results.

Out-of-Sample Forecast	$MAPE_1$	$MAPE_2$	IR (%)
US	0.1710	0.1405	17.87%
CA	0.1986	0.1986	0.03%
GB	0.1393	0.0766	44.96%
DE	0.1296	0.0860	33.63%
FR	0.2147	0.2062	3.94%
IT	0.1575	0.1536	2.46%
AU	0.1009	0.0901	10.65%
JP	0.0260	0.0137	47.13%
IN	0.0461	0.0050	89.18%

Note: $MAPE_1$ and $MAPE_2$ indicate the mean absolute percentage errors of the benchmark model and the EEMD–ARX model, respectively.

actual tourist arrivals and the predicted values. As shown in Figure 5, the shaded areas indicate that the new model significantly outperformed the benchmark model in terms of the reduction of forecasting errors in nine countries.

The results suggested that the proposed methodology is superior to the benchmark model in terms of reducing forecasting errors. The shaded areas indicate our methodology had lower forecasting errors than the benchmark model without decomposition. The superiority of our model with the decomposition process is probably based on the following reasons. First, as demonstrated in the decomposition process, the decomposed IMF3 of the Google search data can capture the turning points and recession periods when tourist arrival data decreased. For the tourist arrival data from the United States, Canada, Great Britain, and France,

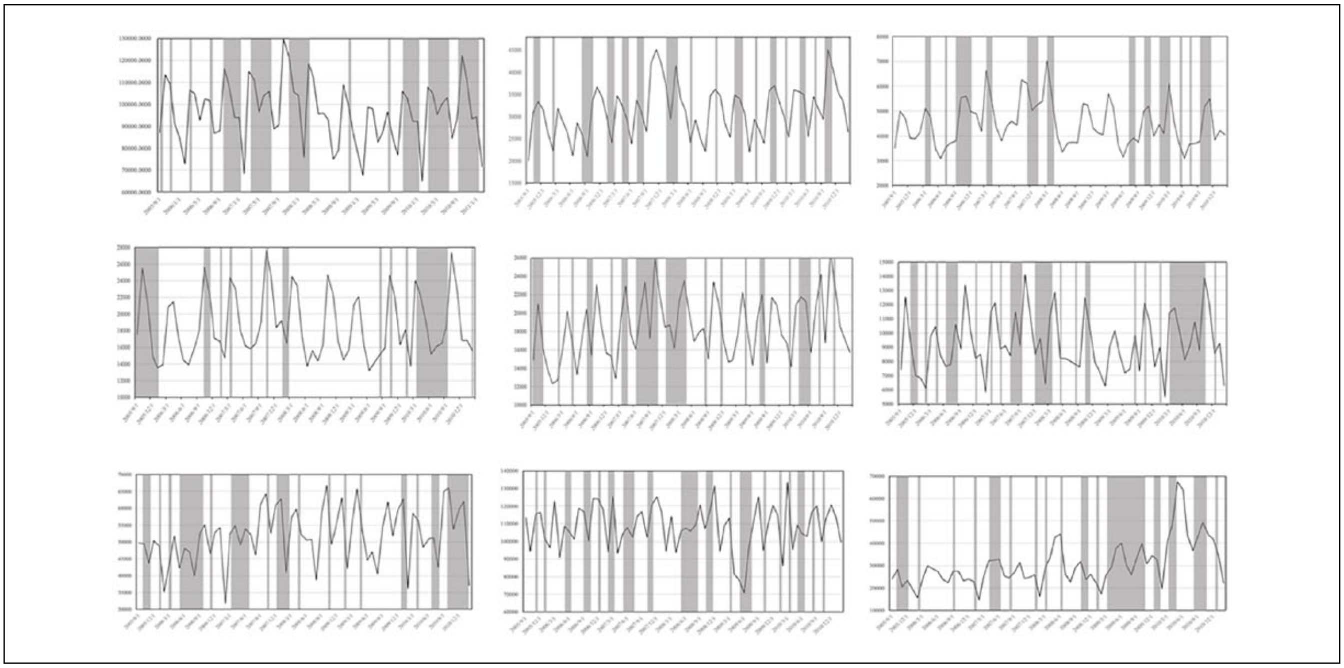

Figure 5. Forecasting results for the US, CA, GB, DE, FR, IT, AU, JP, and IN data sets.

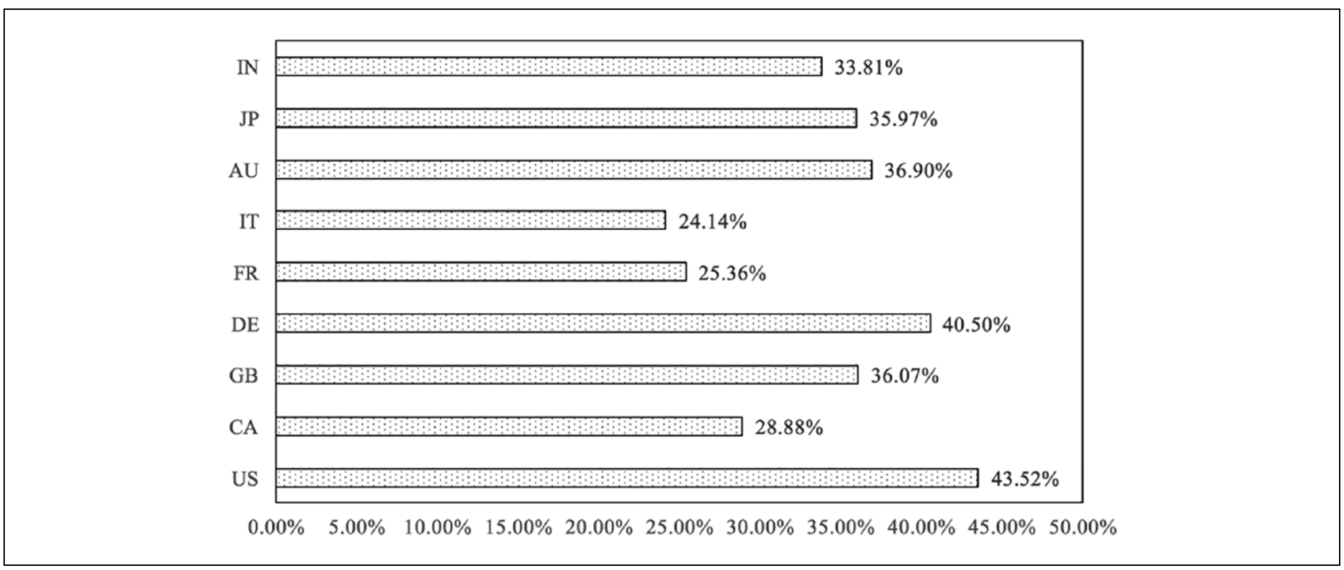

Figure 6. Average increased percentage of out-of-sample forecasting accuracy.

the proposed methodology can accurately predict 27 turning points out of 43, 23 out of 32, 22 out of 36, and 19 out of 29, respectively, compared with the benchmark model. The benchmark model failed to predict these values at turning points, whereas the EEMD–ARX model had the high probability of predicting the values at the turning points accurately. The results indicated the proposed methodology was favorable for predicting directional changes or turning points of tourism demand.

We also computed the average percentage increase in the accuracy of the EEMD–ARX model compared with the benchmark from September 2005 to February 2011. The results are shown in Figure 6. The results showed the EEMD method improved the forecasting accuracy for India, Japan, Australia, Italy, France, Germany, Great Britain, Canada, and the United States by 33.81%, 35.97%, 36.90%, 24.14%, 25.36%, 40.5%, 36.07%, 28.88%, and 43.52%, respectively. Thus, forecasting

Table 5. Improvement Ratios in Different Time Periods.

US	IR (%)	CA	IR (%)	GB	IR (%)
2006.12–2007.1	60.6	2005.10–2005.11	72.0	2006.9–2006.12	19.9
2007.3–2007.5	39.0	2006.8–2006.10	53.6	2007.11–2008.1	9.6
2007.11–2008.3	29.3	2007.7–2007.8	1.4	2009.7–2009.8	85.9
2009.10–2010.1	61.8	2008.1–2008.3	5.1	2009.10–2009.11	84.4
2010.3–2010.7	28.2	2009.3–2009.4	22.1	2010.1–2010.3	40.7
2010.9–2011.11	27.2	2010.9–2010.10	25.9	2010.9–2010.11	49.8
DE	IR (%)	FR	IR (%)	IT	IR (%)
2005.9–2006.1	34.8	2005.9–2005.11	39.5	2005.11–2005.12	12.9
2006.10–2006.11	28.0	2006.8–2006.9	1.4	2006.6–2006.7	1.2
2008.1–2008.2	5.3	2007.6–2007.10	28.9	2007.7–2007.9	21.1
2010.3–2010.9	54.8	2007.12–2008.4	20.9	2007.12–2008.3	12.8
2007.7	89.4	2010.3–2010.5	37.3	2010.3–2010.10	31.0
AU	IR (%)	JP	IR (%)	IN	IR (%)
2005.10–2005.11	9.9	2006.5–2006.6	27.6	2005.10–2005.12	46.8
2006.5–2006.9	45.7	2007.5–2007.7	51.5	2007.4–2007.6	23.2
2007.3–2007.6	33.2	2008.5–2008.8	63.3	2008.10–2008.11	20.7
2007.12–2008.2	8.2	2008.10–2008.11	72.3	2009.3–2009.10	18.6
2009.12–2010.1	80.9	2009.5–2009.7	14.0	2010.8–2010.10	90.4
2010.9–2011.1	41.9	2010.5–2010.6	42.6	2009.12–2010.1	21.8

accuracy was improved by 34% in the prediction of tourist arrivals to Hong Kong from the nine countries.

Table 5 illustrates when exactly the EEMD-based methodology outperformed the benchmark. The results showed that in some representative periods, the proposed methodology achieved a high forecasting accuracy. In Table 5, IR represents the improvement ratio of the EEMD–ARX model compared with the benchmark model during specific period. These periods show the directional change in tourist arrivals. For instance, from December 2006 to January 2007 and from March 2007 to May 2007, tourist arrivals from the United States presented decreasing trends. From August 2006 to October 2006, the tourist arrivals from Canada experienced decreasing and increasing trends, which included the turning point.

The forecasting accuracy was improved significantly in the cases of the tourist arrivals from Japan and India. In particular, forecasting errors were reduced to 63.3% from May 2008 to August 2008 compared with the benchmark model. Forecasting accuracy was improved at 90.4% from August 2010 to October 2010 in the forecasting of tourist arrivals to Hong Kong from India. Figures 7 and 8 present the comparisons between the actual tourist arrivals, the forecast of the benchmark model, and the forecast of the EEMD–ARX model for the tourist arrivals from Japan and India. As indicated in Figure 7, the forecasting accuracy of the EEMD model on October 2008 was higher than the benchmark model by 79%. The MAPE of the EEMD model is 0.054%, whereas the MAPE of the benchmark model is 0.26% on October 2008.

Therefore, adding the components of the search engine data largely improved the overall forecasting accuracy. The EEMD models with Google search data can capture the directional changes in tourism demand. The rationale for

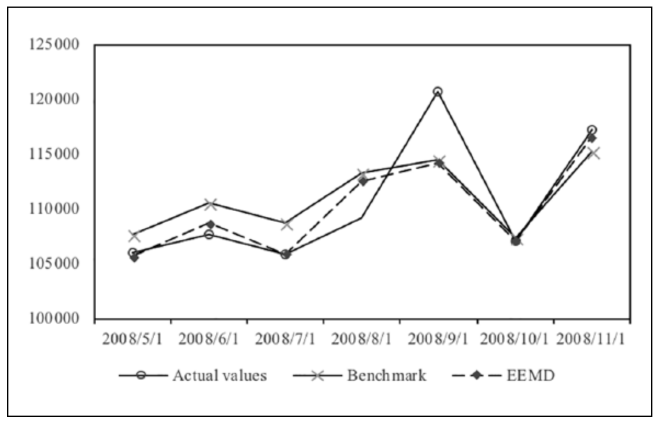

Figure 7. Comparison between the actual tourist arrivals to Hong Kong from Japan and the predictions of the compared models.

this conclusion is explained as follows: the cyclical component extracted using the EEMD method can represent the periodical changes in the search engine data. The cyclical component includes effective information to predict tourism demand accurately. In summary, the results indicated the extracted cycles can help reduce the forecasting errors with respect to the original search engine data. In particular, the proposed methodology can predict tourist arrivals accurately at most turning points.

Concluding Remarks

This study investigated the forecasting ability of the EEMD-based methodology using search engine data in tourism

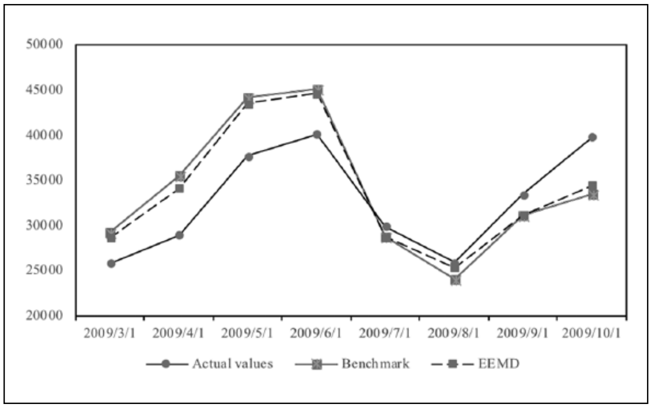

Figure 8. Comparison between the actual tourist arrivals to Hong Kong from India and the predictions of the compared models.

forecasting. By decomposing tourist arrivals to Hong Kong from nine countries, we found similar patterns in the search engine data. The major cycles were first identified and then incorporated into the EEMD–ARX forecasting model. The forecasting results demonstrated the performance of the proposed methodology was largely improved. Compared with the benchmark model in the study of Choi and Varian (2012), our proposed methodology reduced the average forecasting errors by 34% in the prediction of tourist arrivals to Hong Kong from nine countries. The results indicated the proposed EEMD-based methodology using search engine data is useful for tourism forecasting.

This study provided a new perspective on tourism forecasting by exploring the utility of decomposition methods in extracting components from search engine data. The results from the decomposition of search engine data illustrated the intrinsic features of the online data. In particular, similar patterns can be observed from the cycles and trends between the search engine data and tourist arrival data through the individual components. In addition, the proposed methodology can potentially facilitate the identification and forecasting of turning points. This study compared the forecasting ability of the proposed model with the benchmark model and found that the proposed model was useful for the accurate forecasting of most turning points. Focus should be given on the turning points (peaks and valleys) of the tourism market. From this perspective, the research findings can provide support for making decisions in tourism management, especially before and during holiday seasons.

Methodologically, this study employed the data set and extended the forecasting experiments in the study of Choi and Varian (2012), who used Google Trends data to predict tourist arrivals to Hong Kong from nine countries. Their work provided a reliable benchmark. Using this benchmark, we compared our models and their results with those in the existing study and evaluated the forecasting performance of the proposed methodology. Accordingly, our findings are arguably robust because they can be verified. However, the current study still has some limitations. First, the monthly Google Trends data in the study of Choi and Varian (2012) starting from January 2004 to February 2011 cannot be updated without any changes. Given that Google normalizes search engine data, the data will change if the periods are changed. Therefore, we did not revise the data set but used the same one to investigate the performance of the proposed methodology in tourism forecasting in Hong Kong. Second, the number of decomposed IMFs depends on the sample length, which may change with data set changes. However, even with a different number of components, the data series reflecting the cycles can still be extracted. The primary emphasis is to identify the decomposed cycles and incorporate them into the forecasting models. Therefore, the number of decomposed search engine data will not influence the performance of the proposed methodology.

In view of the popularity of massive Internet-generated data, advanced analytics should be used to incorporate them into tourism forecasting. Apart from the EEMD method, nonparametric approaches may be useful in handling search engine data. With these techniques, the sizable online data can be understood and utilized further. The online search engine data and social media data should be analyzed using appropriate approaches to explore influential patterns that affect tourism planning and management.

Author's Note

Xin Li is now affiliated with Collaborative Innovation Center of eTourism, Institute of Tourism, Beijing Union University, Beijing, China.

Acknowledgments

The first author would like to acknowledge the funding from National Natural Science Foundation of China (No. 71601021).

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received the following financial support for the research, authorship, and/or publication of this article: The study was partly supported by a research grant from the National Natural Science Foundation of China (No. 71601021).

References

Akın, M. 2015. "A Novel Approach to Model Selection in Tourism Demand Modeling." *Tourism Management* 48:64–72.

Artola, C., F. Pinto., and P. de Pedraza García. 2015. "Can Internet Searches Forecast Tourism Inflows?" *International Journal of Manpower* 36 (1): 103–16.

Askitas, N., and K. F. Zimmermann. 2009. "Google Econometrics and Unemployment Forecasting." *Applied Economics Quarterly* 55 (2): 107–20.

Assaf, A. G., G. Li, H. Song, and M. G. Tsionas. 2019. "Modeling and Forecasting Regional Tourism Demand Using the Bayesian Global Vector Autoregressive (BGVAR) Model." *Journal of Travel Research* 58 (3): 383–97.

Bangwayo-Skeete, P. F., and R. W. Skeete. 2015. "Can Google Data Improve the Forecasting Performance of Tourist Arrivals? Mixed-data Sampling Approach." *Tourism Management* 46:454–64.

Beneki, C., B. Eeckels, and C. Leon. 2012. "Signal Extraction and Forecasting of the UK Tourism Income Time Series: A Singular Spectrum Analysis Approach." *Journal of Forecasting* 31 (5): 391–400.

Brynjolfsson, E., T. Geva, and S. Reichman. 2016. "Crowd-Squared: Amplifying the Predictive Power of Search Trend Data." *MIS Quarterly* 40 (4): 941–61.

Camacho, M., and M. J. Pacce. 2017. "Forecasting Travellers in Spain with Google's Search Volume Indices." *Tourism Economics*. doi:10.1177/1354816617737227.

Chamberlin, G. 2010. "Googling the Present." *Economic & Labour Market Review* 4 (12): 59–95.

Chen, J. L., G. Li, D. C. Wu, and S. Shen. 2019. "Forecasting Seasonal Tourism Demand Using a Multiseries Structural Time Series Method." *Journal of Travel Research* 58 (1): 92–103.

Chen, C. F., M. C. Lai, and C. C. Yeh. 2012. "Forecasting Tourism Demand based on Empirical Mode Decomposition and Neural Network." *Knowledge-Based Systems* 26:281–87.

Chen, H., R. H. Chiang, and V. C. Storey. 2012. "Business Intelligence and Analytics: From Big Data to Big Impact." *MIS Quarterly* 36 (4): 1165–88.

Cho, V. 2001. "Tourism Forecasting and Its Relationship with Leading Economic Indicators." *Journal of Hospitality & Tourism Research* 25 (4): 399–420.

Choi, H., and H. Varian. 2012. "Predicting the Present with Google Trends." *Economic Record* 88 (Suppl 1): 2–9.

Coshall, J. 2000. "Spectral Analysis of International Tourism Flows." *Annals of Tourism Research* 27 (3): 577–89.

Dergiades, T., E. Mavragani, and B. Pan. 2018. "Google Trends and Tourists' Arrivals: Emerging Biases and Proposed Corrections." *Tourism Management* 66:108–20.

Engelberg, J., and P. Gao. 2011. "In Search of Attention." *Journal of Finance* 66 (5): 1461–99.

Fesenmaier, D. R., Z. Xiang, B. Pan, and R. Law. 2011. "A Framework of Search Engine Use for Travel Planning." *Journal of Travel Research* 50 (6): 587–601.

Fu, Y., J. X. Hao, X. Li, and C. H. Hsu. 2019. "Predictive Accuracy of Sentiment Analytics for Tourism: A Meta-learning Perspective on Chinese Travel News." *Journal of Travel Research* 58 (4): 666–79.

Frechtling, D. 2012. *Forecasting Tourism Demand: Methods and Strategies*. Oxford: Butterworth Heinemann.

Ginsberg, J., M. H. Mohebbi, R. S. Patel, L. Brammer, M. S. Smolinski, and L. Brilliant. 2009. "Detecting Influenza Epidemics Using Search Engine Query Data." *Nature* 457 (7232): 1012–14.

Goel, S., J. M. Hofman, S. Lahaie, D. M. Pennock, and D. J. Watts. 2010. "Predicting Consumer Behavior with Web Search." *Proceedings of the National Academy of Sciences* 107 (41): 17486–90.

Goh, C., and R. Law. 2002. "Modeling and Forecasting Tourism Demand for Arrivals with Stochastic Nonstationary Seasonality and Intervention." *Tourism Management* 23 (5): 499–510.

Gouveia, P. M., and P. M. Rodrigues. 2005. "Dating and Synchronizing Tourism Growth Cycles." *Tourism Economics* 11 (4): 501–15.

Gunter, U., and I. Önder. 2016. "Forecasting City Arrivals with Google Analytics." *Annals of Tourism Research* 61:199–212.

Hamilton, J. D. 1994. *Time Series Analysis*. Princeton, NJ: Princeton University Press.

Huang, N. E., Z. Shen, and S. R. Long, et al.1998. "The Empirical Mode Decomposition Method and the Hilbert Spectrum for Non-stationary Time Series Analysis." *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* 454 (1971): 903–95.

Hassani, H., A. Webster, E. S. Silva, and S. Heravi. 2015. "Forecasting US Tourist Arrivals Using Optimal Singular Spectrum Analysis." *Tourism Management* 46:322–35.

Jackman, M., and S. Naitram. 2015. "Research Note: Nowcasting Tourist Arrivals in Barbados—Just Google It!" *Tourism Economics* 21 (6): 1309–13.

Li, G., K. K. Wong, H. Song, and S. F. Witt. 2006. "Tourism Demand Forecasting: A Time Varying Parameter Error Correction Model." *Journal of Travel Research* 45 (2): 175–85.

Li, G., H. Song, and S. F. Witt. 2005. "Recent Developments in Econometric Modeling and Forecasting." *Journal of Travel Research* 44 (1): 82–99.

Li, X., B. Pan, R. Law, and X. Huang. 2017. "Forecasting Tourism Demand with Composite Search Index." *Tourism Management* 59:57–66.

Li, X., W. Shang, S. Wang, and J. Ma. 2015. "A MIDAS Modelling Framework for Chinese Inflation Index Forecast Incorporating Google Search Data." *Electronic Commerce Research and Applications* 14 (2): 112–25.

Long, W., C. Liu, and H. Song. 2019. "Pooling in Tourism Demand Forecasting." *Journal of Travel Research* 58 (7): 1161–74.

Nguyen, T. L., J. C. Huang, C. C. Chiu, M. H. Shu, and W. R. Tsai. 2013. "Forecasting Model for the International Tourism Demand in Taiwan." Paper Presented at the 2013 International Conference on Technology Innovation and Industrial Management, Phuket, Thailand.

Önder, I., and U. Gunter. 2016. "Forecasting Tourism Demand with Google Trends for A Major European City Destination." *Tourism Analysis* 21 (2/3): 203–20.

Pan, B., C. Wu, and H. Song. 2012. "Forecasting Hotel Room Demand Using Search Engine Data." *Journal of Hospitality and Tourism Technology* 3 (3): 196–210.

Pan, B., and Y. Yang. 2017. "Forecasting Destination Weekly Hotel Occupancy with Big Data." *Journal of Travel Research* 56 (7): 957–70.

Park, S., J. Lee, and W. Song. 2016. "Short-Term Forecasting of Japanese Tourist Inflow to South Korea Using Google Trends Data." *Journal of Travel & Tourism Marketing* 34 (3): 357–68.

Purcell, K., J. Brenner, and L. Rainie. 2012. *Search Engine Use 2012*. Washington, DC: Pew Internet & American Life Project.

Shabri, A. 2016. "A Hybrid of EEMD and LSSVM-PSO Model for Tourist Demand Forecasting." *Indian Journal of Science and Technology* 9 (36).

Song, H., and G. Li. 2008. "Tourism Demand Modelling and Forecasting—A Review of Recent Research." *Tourism Management* 29 (2): 203–20.

Song, H., S. F. Witt, and G. Li. 2008. *The Advanced Econometrics of Tourism Demand*. New York: Routledge.

Vosen, S., and T. Schmidt. 2011. "Forecasting Private Consumption: Survey-Based Indicators vs. Google Trends." *Journal of Forecasting* 30 (6): 565–78.

Vu, H. Q., G. Li, R. Law, and Y. Zhang. 2018. "Tourist Activity Analysis by Leveraging Mobile Social Media Data." *Journal of Travel Research* 57 (7): 883–98.

Wang, C. H. 2004. "Predicting Tourism Demand Using Fuzzy Time Series and Hybrid Grey Theory." *Tourism Management* 25 (3): 367–74.

Witt, S. F., and C. A. Witt. 1995. "Forecasting Tourism Demand: A Review of Empirical Research." *International Journal of Forecasting* 11 (3): 447–75.

Wu, Z., and N. E. Huang. 2009. "Ensemble Empirical Mode Decomposition: A Noise-Assisted Data Analysis Method." *Advances in Adaptive Data Analysis* 1 (1): 1-41.

Xu, X., Y. Qi, and Z. Hua. 2010. "Forecasting Demand of Commodities after Natural Disasters." *Expert Systems with Applications* 37 (6): 4313–17.

Yang, X., B. Pan, J. A. Evans, and B. Lv. 2015. "Forecasting Chinese Tourist Volume with Search Engine Data." *Tourism Management* 46:386–97.

Yang, Y., B. Pan, and H. Song. 2014. "Predicting Hotel Demand Using Destination Marketing Organization's Web Traffic Data." *Journal of Travel Research* 53 (4): 433–47.

Yahya, N. A., R. Samsudin, and A. Shabri. 2017. "Tourism Forecasting Using Hybrid Modified Empirical Mode Decomposition and Neural Network." *International Journal of Advances in Soft Computing & Its Applications* 9 (1): 14–31.

Yu, G., and Z. Schwartz. 2006. "Forecasting Short Time-Series Tourism Demand with Artificial Intelligence Models." *Journal of Travel Research* 45 (2): 194–203.

Yu, L., S. Wang, and K. K. Lai. 2008. "Forecasting Crude Oil Price with an EMD-Based Neural Network Ensemble Learning Paradigm." *Energy Economics* 30 (5): 2623–35.

Zhang, G., J. Wu, B. Pan, J. Li, M. Ma, M. Zhang, and J. Wang. 2017. "Improving Daily Occupancy Forecasting Accuracy for Hotels Based on EEMD-ARIMA Model." *Tourism Economics* 23 (7): 1496–514.

Zhang, X., K. K. Lai, and S. Y. Wang. 2008. "A New Approach for Crude Oil Price Analysis Based on Empirical Mode Decomposition." *Energy Economics* 30 (3): 905–18.

Author Biographies

Xin Li, PhD, is an Assistant Professor at Beijing Union University. Her research interests include big data analytics, tourism forecasting, and business intelligence.

Rob Law, PhD, is a Professor at the School of Hotel and Tourism Management, the Hong Kong Polytechnic University. His research interests are information management and technology applications.