Forecasting

Key words:
Time series
Econometric model
Artificial intelligence
Delphi method
Scenario analysis
Forecasting accuracy

Definition

Forecasting is the practice of using scientific approaches to predict future situations and development trends based on existing information, knowledge and experience. Tourism forecasting normally refers to tourism demand forecasting, or the analysis of current tourism market conditions and past development features. The development of tourism forecasting dates to the 1960s, and significant progress has been achieved in both methodology and practical applications for tourism demand analysis over the past half-century.

Given that most tourism products are 'perishable' (e.g. hotel rooms, flight seats, etc.), accurate forecasting can improve tourism planning and market efficiency and is essential to the interests of stakeholders, including tourism-related business operators, relevant government departments and tourists. Businesses in the tourism industry such as hotels, travel agencies, airlines, tourist attractions and shops, for example, rely heavily on tourism demand, and their profitability is determined by their insights into market trends and their ability to meet market demand based on the forecasts. Tourism forecasting can also provide the government with an important reference for the strategic planning of tourism policies, budgets, infrastructure and so on.

Forecasting methods

Various forecasting methods can be categorised as either quantitative or qualitative. Quantitative methods are normally applied in the forecasting process when there is sufficient measurable information from the past. Qualitative methods are more appropriate when there is little quantifiable information but abundant professional knowledge and experience.

Quantitative methods

Compared to qualitative methods, quantitative methods have increased in development and practice over the past few decades, with forecasting techniques largely based on the principles of mathematics and statistics. This type of method includes time series models, econometric models and artificial intelligence (AI) approaches.

Time series models are also known as non-causal models, as they explain tourism demand by considering historical values together with stochastic disturbance specification. With the advantage of less costly data collection and time-saving model estimation processes, these models are frequently applied in tourism forecasting. Time series models have both basic types and advanced types. The basic types include the autoregressive (AR) model, the moving average (MA) model, the naïve model, the single exponential smoothing (ES) model and the historical average (HA) model. The naïve model is the most popular in tourism forecasting literature due to its easy adoption and reasonable results, especially for short forecasting periods.

The advanced types integrate more features into the models, such as trends, seasonality and special events. Among the various advanced models, the autoregressive integrated moving average (ARIMA) model, also called the Box-Jenkins model (Box & Jenkins, 1976), is dominant. With their distinct modelling flexibility, ARIMA-type models have attracted great interest in recent studies and achieved notable developments in forecasting accuracy. As seasonality is one of the most important features in tourism activity, many advanced models integrate this element, including the seasonal ARIMA (SARIMA) model, the seasonal-naïve model, the Holt-Winters ES model and the basic structural model (BSM). In recent studies, these models have proven to have superior performances in forecasting for some destinations.

Econometric models are called causal models, as they presume a causal relationship between tourism demand and its determinant factors. Key determinants include tourists' income level, tourism prices between one destination and its competitors, and the exchange rate between a source market and a destination. Other factors influencing tourism demand include marketing, one-off events, political and economic situations and climates. In current research, the causal relationship is estimated by a regression model after variable selection using the general-to-specific (GETS) approach. Demand forecasts can then be generated from the model, with the estimated causal relationship and the forecast values of the selected determinants (Witt & Witt, 1995).

Econometric models can be divided into single equation models and multiple equation models. In the single equation category, the most basic model is the single static regression (SR) model, which was widely used in early studies and is now regarded as the benchmark for evaluating forecasting performance. Based on the SR model, the distributed lag (DL) model, the autoregressive distributed lag model (ADLM) and the error collection model (ECM) all account for the intertemporal relationships between tourism demand and its influencing factors.

The DL model however, is less advanced and lacks generality compared to the ADLM, so it is mainly applied as a forecasting benchmark. The ECM not only considers the long-term relationships between tourism demand and its determinants as the ADLM does, but also incorporates a short-run mechanism to correct the variables in terms of long-run equilibrium. Both the ADLM and the ECM occupy important positions in tourism forecasting, and their overall performances are outstanding.

Time series models can also integrate causal relationships by introducing exogenous variables (X), such as the ARX, ARIMAX and SARIMAX models. Additionally, the structural time series model (STSM) also originates from this type of extension by introducing explanatory variables into the BSM. The multiple equation models consider the interdependency between different time series or tourism demand equations. The vector autoregressive (VAR) model and the vector ECM (VECM) capture interdependencies between multiple time series, but the predictive results are not significant. The Bayesian VAR (BVAR) model was developed by Wong et al. (2006) to improve the performance of the VAR model by including Bayesian priors. The VAR model can also be applied to global VAR (GVAR) and Bayesian GVAR (BGVAR) models. The almost ideal demand system (AIDS) (Deaton & Muellbauer, 1980) captures the interdependency between multiple demand equations and is usually applied to forecast tourism demand as measured in market share. Panel data regression (PDR) is also a multiple equation method integrating both the intertemporal and cross-sectional features of the data, but it is rarely applied in recent studies.

As fast-growing forecasting methods, AI models are able to explain non-linear quantified information even without prior knowledge or assumptions about the relationships between the forecast variable and its explanatory variables. Among several AI techniques that have been applied in tourism demand analysis, the artificial neural network (ANN) models are the most frequently used, as they provide an outstanding performance in terms of processing imperfect data and estimating a wide range of non-linear relationships (Song et al., 2019). Previous studies indicate that ANNs perform best when making short-term forecasts or processing imperfect time series data, but they may be outperformed by the time series models if the data are pre-processed. However, ANNs have been criticised by researchers, as there is a ‘black box’ or uninterpretable hidden layer between inputs and outputs, and the results lack theoretical support (Zhang et al., 1998). Despite their methodological limitations, applications of AI techniques in tourism forecasting are on the rise. Other popular AI techniques include the support vector machine (SVM), the genetic algorithm, the grey theory, the fuzzy time series and the rough sets theory.

Quantitative Methods	Representative Models	Advantages
Time series	Basic: AR, MA, naïve, single ES, HA	Less costly data collection and time-saving model estimation process
	Advanced: ARIMA, SARIMA, seasonal-naïve, Holt-Winters ES, BSM	
Econometrics	Single equation: SR, DL, ADLM, ECM, ARX, ARIMAX, SARIMAX, STSM	Establishing the causality between tourism demand and its determinant factors
	Multiple equation: VAR, BVAR, GVAR, BGVAR, VECM, AIDS, PDR	
Artificial intelligence	ANN, SVM, fuzzy time series, rough set theory, grey theory	Explaining non-linear information even without prior knowledge about the relationships between the forecast value and its explanatory variables

Table 1 Quantitative methods in tourism demand forecasting

Qualitative methods

Qualitative methods, also called judgemental methods, mainly rely on experts' knowledge, experience and insight. One of the most popular qualitative methods is the Delphi model, which invites anonymous experts to forward their opinions and redeclare them after feedback from other experts until the panel reaches a consensus. The Delphi model is mainly used for long-term tourism forecasting, particularly when knowledge of the forecasting variable is insufficient. However, this approach has been criticised for its high dependency on biased subjective opinions, and it is difficult to test its accuracy. Another important qualitative technique is scenario construction. Experts are required to analyse, discuss and achieve consensus on the forecasts in each selected scenario (Moutinho & Witt, 1995). By setting and defining the essential conditions of future scenarios, the limitations of the Delphi model may be overcome to a certain extent.

Forecasting accuracy

Factors affecting forecasting accuracy include forecasting horizons, data-generating processes, estimation methods and model specification. Forecasting accuracy can be evaluated by the magnitude of forecasting error or the magnitude of turning point error. The former is dominant in tourism studies and is commonly calculated using the mean absolute percentage error (MAPE) or the root mean square percentage error (RMSPE), whereas the latter is strategically important for industrial management and can be quantified by measuring the percentage of correctly forecasted moving directions (Song et al., 2013).

Challenges & critical aspects

As the forecasting process is primarily based on retrospective captures of the development rules of economic and social activities, forecasting can achieve higher accuracy when the macroenvironment is evolving stably and regularly. That said, the effectiveness of forecasting can become quite limited if unexpected events disorder the market, such as natural (e.g. earthquake, volcanic eruption and forest fire) or man-made disasters (e.g. terrorist attack, financial crisis and epidemic disease).

A number of studies have discussed the influences of these events on tourism forecasting and proposed valid methods to capture the changing patterns. However, it is still unrealistic for researchers to foresee the occurrences of these incidents. What forecasting analysis can do is make ex-post adjustments to the original forecasts. Like the SARS epidemic in 2003 and the global financial crisis in 2008, the COVID-19 pandemic is bringing unprecedented challenges to tourism forecasting. Since the outbreak of the COVID-19 in early 2020, many adjusted forecasts of tourism demand have been published, but the results are revised continually due to the rapid evolution of the epidemic. Scenario analysis and other judgemental approaches are being used more widely than ever to help discern the inconceivable future and make reasonable estimations wherever possible.

Opportunities & future developments

Point forecasts have dominated tourism forecasting studies over the past decades. More recently, however, researchers are paying more attention to interval forecasts, which provide corresponding prediction intervals given designated probabilities. Taking future volatility into account, interval forecasts have a higher reference value for stakeholders to formulate policies and plans.

Although tourism forecasting methods have improved significantly and provide better accuracy performance, no single model can consistently outperform the others. Recent studies tend to integrate quantitative and qualitative methods to obtain higher accuracy. Most of the integrations use quantitative models to estimate forecast values, then use qualitative approaches for any necessary adjustments. The integration of quantitative and qualitative methods combines the unique strengths from each and is particularly superior when both sufficient data and experts' knowledge are available. Accordingly, hybrid models and combined models are emerging. The former refers to a new single model formed from the elements from other models, whereas the latter represents the combination of forecast results estimated by different models (Song et al., 2019).

An increase in the use of AI-based methods is also leading to new directions in research. In addition to forecasting applications as independent models, they are being more widely used in association with other quantitative methods.

Fast-growing big data resources are optimising the data selection process. Researchers are able to dig more deeply and explore new variables to refine the forecasting models. One such example is the usage of Internet data (e.g., Google Trends) as exogenous variables to improve time series models.

References

Box, G. E. P., & Jenkins, G. M. (1976). *Time series analysis: Forecasting and control*. Holden-Day.

Deaton, A., & Muellbauer, J. (1980). An almost ideal demand system. *The American Economic Review, 70*(3), 312-326.

Moutinho, L., & Witt, S. F. (1995). Forecasting the tourism environment using a consensus approach. *Journal of Travel Research, 33*(4), 46-50.

Song, H., Qiu, R.T.R., & Park, J. (2019). A review of research on tourism demand forecasting. *Annals of Tourism Research, 75*(2019), 338-362.

Song, H., Smeral, E., Li, G., & Chen, J. L. (2013). Tourism forecasting using econometric models. In C. Costa, E. Panyik, & D. Buhalis (Eds.), *European tourism planning and organisation* (pp. 289-309). Chanel View Publications.

Witt, S. F., & Witt, C. A. (1995). Forecasting tourism demand: A review of empirical research. *International Journal of Forecasting, 11*(3), 447-475.

Wong, K. K., Song, H., & Chon, K. S. (2006). Bayesian models for tourism demand forecasting. *Tourism Management, 27*(5), 773-780.

Zhang, G., Patuwo, B. E., & Hu, M. Y. (1998). Forecasting with artificial neural networks: The state of the art. *International Journal of Forecasting, 14*(1), 35-62.

Authors

Haiyan Song, PhD, School of Hotel and Tourism Management, The Hong Kong Polytechnic University

Gabrielle Lin, MSc, School of Hotel and Tourism Management, The Hong Kong Polytechnic University