Measuring tourism demand nowcasting performance using a monotonicity test

Han Liu

Center for Quantitative Economics and School of Business and Management, Jilin University, Changchun, 130012, P. R. China
Address: Qianjin Avenue, #2699, Gaoxin District, Changchun, 130012. China.
Tel: +86-13578656976; Email:hanliu@jlu.edu.cn

Yongjing Wang

School of Business and Management, Jilin University, Changchun, 130012, P. R. China
Address: Qianjin Avenue, #2699, Gaoxin District, Changchun, 130012. China.
Tel: +86-18844098155; Email: wyj201805@foxmail.com

Haiyan Song

Shenzhen Research Institute, School of Hotel and Tourism Management, The Hong Kong Polytechnic University, Hong Kong SAR
Adresss: 17 Science Museum Road, TST East, Kowloon, Hong Kong
Tel: +852-34002286; Email: haiyan.song@polyu.edu.hk

Ying Liu*

School of Business and Management, Jilin University, Changchun, 130012, PR China
Address: Qianjin Avenue, #2699, Gaoxin District, Changchun, 130012. P. R. China.
Tel: +86-17390940897; Email: liuyingjlu@foxmail.com

Acknowledgment

The authors would like to acknowledge the financial support of the Natural Science Foundation of China (Grant No.: 72004077), Hong Kong Research Grant Committee (Grant No.: PolyU - 15502120).

***Corresponding Author:**
Ying Liu, School of Business and Management, Jilin University, Changchun, 130012, PR China
Address: Qianjin Avenue, #2699, Gaoxin District, Changchun, 130012. P. R. China.
Tel: +86-17390940897; Email: liuyingjlu@foxmail.com

Measuring tourism demand nowcasting performance using a monotonicity test

Abstract

Tourism demand nowcasting is generally carried out using econometric models that incorporate either macroeconomic variables or search query data as explanatory variables. Nowcasting model accuracy is normally evaluated by traditional loss functions. This study proposes a novel statistical method, the monotonicity test, to assess whether the nowcasting errors obtained from the OLS, generalised dynamic factor model (GDFM) and generalised dynamic factor model combined with mixed data sampling (GDFM-MIDAS) model are monotonically decreasing when new data on explanatory variables become available, based on the mixed frequency data between January 1, 2011 and December 31, 2019. The results of the empirical analysis show that nowcasts generated results based on two data sources combined are superior to that based on a single data source. Compared with traditional loss functions, the monotonicity test leads to a more objective and convincing nowcasting model performance. This study is the first attempt to evaluate tourism demand nowcasting performance using a monotonicity test.

Keywords: Tourism demand; nowcasting; monotonicity test; mixed frequency data.

Introduction

The development of tourism in Greater China, which encompasses Hong Kong, Macau, and Taiwan along with the mainland of China, has contributed greatly to the tourism industry across Asia and the wider world (Li, 2009). Factors that have spurred tourism development within Greater China include geographical proximity, cultural similarities, rich natural and cultural resources, diverse cuisines, and minimal language barriers. In addition, the opening of the Hong Kong–Zhuhai–Macau bridge in 2018 further stimulated travel between the mainland of China and the two special administrative regions. As a result, travel within Greater China is now more cost-effective for Chinese mainland residents, for whom Hong Kong, Macau, and Taiwan have become preferred tourism destinations. Against this backdrop, scholars are gradually paying more attention to tourism demand forecasting in Greater China (Li, 2009). This study adds an additional dimension to this line of research.

The vital role that the tourism industry plays in economic development and social and cultural exchange is widely recognised (Wan and Song, 2018). Tourism destinations routinely use scientific methods to market and manage tourists' experiences to increase the appeal of tourism products and services (Song and Li, 2008). Tourism is considered a key engine of economic growth (Jiao et al., 2020). According to the United Nations World Tourism Organization (UNWTO, 2020), the growth rate of international tourism exceeded the growth rate of the global economy in 2019. In the same year, tourism contributed around 10% to global gross domestic product (GDP) and accounted for a similar percentage of jobs created globally (WTTC, 2019). For destinations heavily reliant on the tourism industry, accurate forecasting of future tourism demand has become increasingly important for developing effective strategies and policies (Kulshrestha et al., 2020). In addition, tourism-related organisations, such as marketing agencies, airlines, and transportation providers, have a great interest in monitoring real-time tourism demand fluctuations, known as nowcasting (Hirashima et al., 2017). Although tourism demand nowcasting has received increasing attention in recent years, research on ways to assess the accuracy of nowcasting methods has lagged.

Broadly, nowcasting aims to predict economic activities in the immediate future (e.g., the days and weeks ahead). Therefore, nowcasting could also be classified as short-term forecasting (Castle et al., 2009). In economics, nowcasting is used to deal with delayed data publication (Jackman and Naitram, 2015). Given the importance of nowcasting, the accuracy of nowcasting techniques deserves closer attention. Similar to forecasting accuracy, nowcasting accuracy can be measured with traditional loss functions such as the root mean squared error (RMSE), the mean absolute percentage error (MAPE), the mean absolute error (MAE), the mean squared forecast error (MSFE), the root mean squared percentage error (RMSPE), and Theil's U statistic (Li et al., 2005). However, the effectiveness of these measures declines as high-frequency data is continuously added to the recursive model estimation process, especially when the nowcasting ranges are shorter than the official data publication frequencies (Patton and Timmermann, 2012; Fosten and Gutknecht, 2020). Therefore, this study proposes using a monotonicity test to evaluate the performance of nowcasting models. While such a test has been shown to be very useful in evaluating the performance of GDP nowcasting models (Bańbura and Modugno, 2014; Knotek and Zaman, 2017; Bragoli and Fosten, 2018), there has been no attempt to evaluate tourism demand nowcasting performance using a monotonicity test. This study makes an important contribution for the tourism demand nowcasting literature to fill this gap.

Another contribution of this study is that the nowcasting model specification used includes high-frequency explanatory variables based on existing econometric models used in tourism demand forecasting. Tourism demand research has generally used macroeconomic variables or search query data as explanatory variables (Höpken et al., 2019; Antolini and Grassini, 2019), but few studies have combined these two data types in modelling and nowcasting tourism demand. Studies (Pan et al., 2012; Wen et al., 2019) have shown that there are drawbacks to forecasting tourism demand solely using macroeconomic variables. Indeed, as the digital transformation increasingly takes hold, real-time search query data may better explain people's intention to travel. Thus, adding search query data or online big data as explanatory variables may substantially improve the nowcasting model's performance compared with traditional univariate time series models (Hirashima et al., 2017). This study constructed four nowcasting models with two different data sources, macroeconomic variables and the Baidu index, to generate the nowcasting results of tourism demand from the mainland of China to three destinations. The purpose is to examine whether the performance of nowcasting models may improve as new information is added to the model estimation. In addition, we use a mixed frequency time series model to nowcast tourist arrivals based on search query data and macroeconomic variables.

The rest of the study is organised as follows. Section 2 reviews the literature and describes the methods used in tourism demand nowcasting research. Section 3 introduces the nowcasting models, process, evaluation and data. Section 4 presents the empirical results, with a particular focus on the nowcasting results and their accuracy based on the monotonicity test. The final section summarises the conclusions.

Literature review

Tourism, an economic activity that generates foreign exchange and creates jobs, contributes substantially to the economies of many countries (Brida et al., 2020; Yang et al., 2015). Governments and private firms are interested in accurate tourism demand forecasts, which are essential for effective planning and development of needed marketing and infrastructure (Song and Li, 2008). In the short run, the accurate nowcasting of tourism demand is also vital for enhancing business operations at mass or multiproduct tourism destinations (Emili et al., 2020) and monitoring the effectiveness of ongoing tourism-related policies (Castle et al., 2009; Jackman and Naitram, 2015).

Tourism demand nowcasting

Most published studies on tourism demand modelling and forecasting using econometric approaches have had two objectives: discovering the relationship between tourism demand and its influencing factors and generating accurate forecasts based on the relationship models estimated (Song et al., 2019). Empirical evidence has shown that macroeconomic variables, such as the source market GDP, the tourism prices in the destination relative to that of the source markets, and the substitute prices of competing destinations have an important influence on tourism demand. These explanatory variables also improve the forecasting performance of tourism demand models (Song et al., 2012; Wu et al., 2017).

Tourism nowcasting, a special form of tourism forecasting, involves modelling the immediate past and current dynamics in tourism demand and predicting the near future demand using the model established (Antolini and Grassini, 2019). Nowcasting is widely used in macroeconomic forecasts because it effectively deals with the problem caused by the time lag in statistical data releases.

Studies have generally assessed the accuracy of tourism demand forecasting and nowcasting using traditional loss functions, such as the RMSE, the MAE, the MAPE (Song et al., 2009b; Gunter and Önder, 2015; Jackman and Naitram, 2015), and the direction of change (DC) (Hassani et al., 2013; Silva et al., 2017).

Traditionally, tourism demand forecasting models are estimated using macroeconomic variables. With the growth of the digital economy, however, such traditional economic variables cannot adequately reflect real-time changes in tourism demand (Feng et al., 2019). Furthermore, due to the slow release of macroeconomic data, models using official statistics are unable to capture real-time behavioural changes in tourism demand (Forni et al., 2000; Bangwayo-Skeete and Skeete, 2015). Due to these problems, researchers have begun to model and forecast tourism demand using search query data, which are normally high frequency and real-time in nature (Valdivia and Monge-Corella (2010); Carriere-Swallow and Labbe (2013).

Incorporating search query data into univariate time series models can improve the forecasting accuracy of the target variables. For instance, Pan et al. (2012) showed that introducing search query data into an ARIMA model (ARIMAX) to forecast the demand for hotel rooms resulted in superior model performance relative to a time series model without search query data. Similarly, Yang et al. (2015) empirically demonstrated that using search query data to forecast tourist arrivals in Hainan Province helped to improve forecasting performance substantially. Pan et al. (2017) combined search queries and Internet traffic data to forecast weekly hotel demand; the forecasts of the ARIMAX model that included the combined online data were more accurate than those of the ARIMA model. Similar conclusions were reached by Li et al. (2017), Wen et al. (2019), and Hu et al. (2021). However, there is also evidence that search query data cannot always produce better predictions, as forecasting performance depends on the population targeted by specific search engines (Yang et al., 2015; Li et al., 2020; Hu et al., 2021). For example, the Baidu index is more useful than Google search queries for forecasting tourism demand from Chinese tourists (Volchek et al., 2019).

Search query data can reflect the latest consumer intention, especially those related to new socio-economic phenomena. This information can supplement official statistics and assist business decisions (Huang and Hao, 2021). Thus, the inclusion of search query data in tourism demand forecasting models that mainly include macroeconomic variables can improve model forecasting performance (Siliverstovs and Wochner, 2018), with such data being particularly useful in understanding real-time behavioural changes (Antolini and Grassini, 2019). However, few studies have focused on combining macroeconomic variables and search query data to forecast tourism demand.

Studies have added search query data to basic time series models (Bangwayo-Skeete and Skeete, 2015; Dimpfl and Langen, 2018). This study, in contrast, includes search query data in an econometric model that contains traditional economic variables to analyse whether tourism demand nowcasting performance can be improved.

Tourism demand forecasting evaluation

The most common indexes using accuracy measures in tourism demand forecasting are the RMSE, MAE, MAPE, and DC (Jiao and Chen, 2019; Silva et al., 2017), with the RMSE and MAE being the most widely used forecast error measures. Between these two measures, the MAE is more sensitive to small deviations from zero. It is less sensitive to large deviations because it is not calculated based on the squared loss values. Some scholars have used Theil's U statistic to measure forecast accuracy

(Song et al., 2009b). Another method sometimes used to evaluate model performance is the RMSPE (Gunter and Önder, 2015). Generally, however, the RMSE, MAE, and MAPE are the most frequently used measures.

As just mentioned, if a model's forecasting performance is measured by the squared forecast errors, the measure will be very sensitive to large changes in forecast errors (Gunter and Önder, 2015). Therefore, the evaluation of a model's forecasting performance depends largely on the measure used in the assessment. Meanwhile, it is still unknown whether a tourism demand model's forecasting accuracy can be monotonously improved as the sample size of explanatory variables expands over time. This study seeks to address this issue.

Specifically, this study uses a monotonicity test to evaluate the nowcasting performance of tourism demand models based on mixed frequency tourism demand data collected in Greater China. Nowcasting monotonicity test has been used as an evaluative criterion by policy-making institutions, such as the Atlanta Fed. It has also been used in empirical papers dating back to Giannone et al. (2008), who applied the uncertainty measurement to indicate nowcasting performance. Fosten and Gutknecht (2020) proposed a formal and robust test for nowcasting monotonicity based on the moment inequality procedure of Chernozhukov et al. (2019). The monotonicity test, which can be applied in general settings, represents the first rigorous procedure to assess nowcasting performance. Its purpose is to determine whether the accuracy of a nowcasting model using big data shows a monotonic decreasing trend as new information is continuously included in the model estimation. The monotonicity test used in our study is developed based on the multiple moment inequality procedure proposed by Corradi and Swanson (2014). This method is superior to the forecasting accuracy measures used in previous studies. Rather than using a formal monotonicity test, Marcellino et al. (2016) simply used the trend of the conventional MSFE over the forecasting period to determine whether there was a decline in forecast accuracy. Furthermore, compared with the monotonicity tests used in other studies, the one used in this study is more objective and reliable. For example, this study extends the univariate monotonicity test of Bańbura et al. (2013) to multiple moment inequalities to evaluate the performance of tourism demand nowcasting models. To the best of our knowledge, our study is the first to formally test whether tourism demand nowcasting accuracy is monotonic.

In summary, the contributions of this study to the literature are presented. First, although tourism demand nowcasting is attracting growing interest from tourism scholars, researchers continue to use conventional forecasting performance measures, such as the RMSE, MAE, and MAPE, to evaluate the nowcasting performance of models. Evidence has shown that these forecasting performance measures become less sensitive as new observations regarding explanatory variables are continuously updated (Gunter and Önder, 2015). Second, studies on tourism demand forecasting have failed to include macroeconomic variables and search query data simultaneously as explanatory variables in the models. Therefore, this study is the first to use mixed frequency data in model specification and nowcasting with a novel performance evaluation method.

Methodology and data

Tourism demand nowcasting

Based on Song and Romilly (2000), the most important factors affecting tourism demand are the destination of the own price, the substitutes price, and the income of consumers. The following mathematical function is used to describe the relationship between tourism demand and its influencing factors:

$$VA_{j,t} = AP_{j,t}^{\beta_p} Y_{j,t}^{\beta_y} P_{j,s,t}^{\beta_{p_s}} e_{j,t}, \tag{1}$$

where $VA_{j,t}$ is tourism demand, measured by visitor arrivals from the mainland of China to destination j (j representing Hong Kong, Macau, and Taiwan, respectively) at time *t*; $P_{j,t}$ is the destination of the own price of tourism in Hong Kong/Macau/Taiwan at time *t*, adjusted by the respective exchange rates; $Y_{j,t}$ is the income level of Chinese mainland residents at time *t*, measured by China's IP (industrial production index); $P_{j,s,t}$ is the substitute price of tourism in the competing destination(s) at time *t*. $e_{j,t}$ is a residual item capturing the influence of all other factors that are not included in the model.

There are two purposes for using Equation (1). First, the power function better reflects the relationship between tourism demand and its influencing factors. Second, the power function can easily be transformed into a linear relationship through log transformation (see Equation (2)), and this is easy to estimate existing estimators such as ordinary least squares (OLS). In addition, the coefficients of Equation (2), apart from the constant, are demand elasticities.

$$\ln VA_{j,t} = \beta_0 + \beta_p \ln P_{j,t} + \beta_y \ln Y_{j,t} + \beta_{p_s} \ln P_{j,s,t} + u_{j,t}, \tag{2}$$

where $\beta_0 = \ln A$, $u_{j,t} = \ln e_{j,t}$,and $\beta_p, \beta_y, \beta_{p_s}$ which represents the estimated coefficient after taking the logarithm of Equation 1, they are the price, income, and substitution price elasticities, respectively. The signs of the coefficients are $\beta_p < 0$ and $\beta_y > 0$, $\beta_{p_s} > 0$.

Research has shown that the generalised dynamic factor model (GDFM) can produce superior nowcasting performance in tourism demand nowcast, which forms the basis of this study. Meanwhile, a mixed data sampling (MIDAS) model is used to deal with data having different frequencies. Hence, the model used for the empirical analysis in this study is known as the GDFM-MIDAS model. To use the GDFM-MIDAS model to nowcast tourism demand, we extract the optimal number of factors firstly from the available time series by the GDFM-MIDAS model. Then, these extracted factors are included in the GDFM-MIDAS model as explanatory variables for tourism demand nowcasting.

The GDFM-MIDAS model

This section introduces the GDFM-MIDAS model, focusing on the model specification and estimation. Nowcasting performance evaluation methods are also introduced. To better understand the GDFM-MIDAS model, it is essential to define the MIDAS model firstly.

1. The MIDAS model

The MIDAS model allows high-frequency explanatory variables to explain low-frequency dependent variables. Its general form can be written as follows:

$$y_t = \beta_0 + \beta_1 W(L^{1/m};\theta) x_t^{(m)} + \varepsilon_t^{(m)}, \tag{3}$$

where y_t is the low-frequency dependent variable, $x_t^{(m)}$ is a high-frequency explanatory variable, and *m* is the observed *m* times in the same period. For example, *m*=31 means the nowcast made when there has a new data update for the 31st time in the current month. β_0, β_1 indicate the estimated coefficient obtained by the MIDAS model. All of the MIDAS models in this study are estimated using the two-parameter exponential Almon polynomial, the specific form of which can be written as follows:

$$\omega(k,\theta)=\frac{e^{(\theta_1 k+\theta_1 k^2)}}{\sum_{k=1}^{K} e^{(\theta_1 k+\theta_1 k^2)}}. \tag{4}$$

The estimated MIDAS model is obtained by minimising the residuals in Equation (3):

$$\left[\hat{\beta},\hat{\theta}\right]=\underset{\beta,\theta}{\arg\min}\sum_{t}(y_t-\beta_0-\beta_1 W(L^{1/m};\theta)x_t^{(m)})^2. \tag{5}$$

2. The GDFM model

The factor model is used to define and measure intelligence. Factor analysis aims to describe the correlation between variables using a small number of potential and unobservable factors (Li. et al., 2020).

Forni et al. (2000) proposed the GDFM by extending dynamic factor models. They argued that "dynamic" and "approximate" are two important characteristics of a factor model to solve time series data. First, analysing time-series data is a typical dynamic problem. The model must allow heterogeneity to be a cross-sectional correlation for other cross-sectional data. The orthogonality assumption of heterogeneity is unrealistic for most typical dynamic problems. Therefore, the GDFM is better suited to tourism demand forecasting. It consists of two parts: its common component $\chi_{i,t}$ and an idiosyncratic component $\xi_{i,t}$. Therefore, for the observed variables $\{\mathrm{X}_{i,t}, i=1,...,n; t=1,...,T\}$, the model is expressed as follows:

$$\mathrm{X}_{i,t}=\chi_{i,t}+\xi_{i,t} \tag{6}$$

$$\chi_{i,t}=b_{i,1}(L)f_{1,t}+b_{i,2}(L)f_{2,t}+...+b_{i,q}(L)f_{q,t}, \tag{7}$$

where the common components $\chi_{i,t}$ are driven by a q-dimensional vector of common factors $(f_{1,t}, f_{2,t},..., f_{q,t})$. $b_{i,k}(L)=\sum_{l=1}^{\infty} b_{i,k,l}L^l, k=1,...,q$ is the set of time-varying factor loadings, L is the lag operator, and q indicates the number of unobserved commonly dynamic factors. The number q of common shocks throughout has been regarded as a known factor, and we assume it to be constant over time. According to Barigozzi et al. (2021), the number of factors has to be estimated from the observations. This study uses the criterion proposed by Hallin and Liška (2007) to determine these factors. Forni et al. (2000) suggested that q is determined through the variance contribution of each component.

One feature of the GDFM model is the estimation of common factors. For example, we aim at nowcasting tourism demand using hundreds of Baidu indexes. In the GDFM framework, these observed variables can partly be explained by common unobserved factors, which are noted as common components $\chi_{i,t}$. The common components are useful information that can be used to explain tourism demand and aggregate the relevant index.

Based on Song and Romilly (2000), we can specify the three commonly used traditional variables, which are the price of the substitute destination, the price of tourism in the destination, and the per capita income in the country of origin to build Model 1. The macroeconomic factors and the factors extracted from the Baidu index are added recursively to the model. In this way, the macroeconomic and Baidu index factor model can model and nowcast tourism demand for the specific destination under consideration. The nowcasting results are then evaluated by the monotonicity test. Specifically, the GDFM-MIDAS tourism demand model can be written as follows:

$$y_t = \beta_0 + \sum_{i=1}^{n} \beta_i W_i(L^{1/m};\theta_i) \mathrm{X}_{i,t}^{(m)} + \varepsilon_t^{(m)}, \tag{8}$$

where $\beta_0, \beta_i (i=1,2,..,n)$ indicate the estimated coefficient obtained by the GDFM-MIDAS model.

Nowcasting process

Most published studies on tourism demand forecasting using search query data (Baidu index) have been based on univariate time series models, with search queries as explanatory variables (Pan and Yang, 2017; Volchek et al., 2019). This study, however, starts with Model 1 and gradually adds other macroeconomic factors and Baidu index factors.

First, based on Song et al. (2011), Chatziantoniou et al. (2016), Wu et al. (2017), and Nor et al. (2018), this study incorporates the other macroeconomic variables—i.e., economic policy uncertainty (*EPU*), consumer price differentials (*CPDs*), consumer confidence index (*CCI*), consumer price index (*CPI*), and the logarithmic form of the lag value of visitor arrivals (VA_{lags}) into Model 1 to construct Model 2. The purpose of this process is to analyse whether adding different variables contributes to the nowcasting performance of the tourism demand model.

Second, using the GDFM-MIDAS model (Li et al., 2017), we incorporate daily Baidu index factors into Model 1 to create Model 3.

Finally, based on Model 1, the macroeconomic factors and Baidu index factors are both added to form Model 4.

Table 1 shows the specific forms of the four models (Models 1 through 4), and Figure 1 describes the process of the current research.

[Insert Table 1 about here]

[Insert Figure 1 about here]

Based on the above comparison model, this study first analyses the nowcasting performance of models, and then generates the monotonicity test results based on the tourism demand nowcasting results from the mainland of China to three destinations. Specifically, the first step is to divide the total data sample into two parts: (1) the fitting set: data covering the period from January 2011 to December 2018 are used for in-sample estimation; (2) the nowcasting set: data covering the period from January 2019 to December 2019 are used to generate the nowcasting results. Then, the second step is to evaluate the nowcasting performance of different models constructed using different data sources.

Nowcasting evaluation

This section adopts the traditional approach and monotonicity test to evaluate the tourism demand nowcasting accuracy.

Traditional approach

Based on Models 1 through 4, this study carries out tourism demand nowcasting, and the results are evaluated using the following three measurements.

$$RMSE = \sqrt{\frac{1}{T-R}\sum_{t=1}^{T-R}(\hat{y}_t - y_t)^2}, \quad (9)$$

$$MAE = \frac{1}{T-R}\sum_{t=1}^{T-R}|\hat{y}_t - y_t|, \quad (10)$$

$$MAPE = 100 \times \sum_{t=1}^{T-R}|1-\hat{y}_t/y_t|/(T-R), \quad (11)$$

where the length of the full sample is *T*, and we split the length of the in-sample is *R*, and the length of the out-of-sample is (*T*-*R*), y_t represents the actual values of period

t, and $\hat{y}_t$ represents the forecast values of period t. The smaller value of these indicators, the smaller gap between the nowcasting value $\hat{y}_t$ and the actual value y_t, which means more accurate nowcasting results were obtained.

Monotonicity test

In general, most studies have used evaluation methods such as the RMSE to test whether the forecasting performance of a given method gradually improves as data are updated and added (Patton and Timmermann, 2012; Marcellino et al., 2016). This study applies a test proposed by Fosten and Gutknecht (2020) to determine whether big data nowcasting methods, which have become an important tool for many public and private institutions, monotonically improve as new information becomes available. Corradi and Swanson (2014) propose the monotonicity test, which is a formal and rigorous method used to evaluate forecasting performance based on the multiple moment inequality. The number of forecasts approaches infinity, which means that the number of moment inequalities tested can do the same; hence, this model is suitable for testing the results of factor models and other forecasts using high-dimensional data. The purpose of this study is to make a monotonic assessment of the forecast results of the explained variables y_t for the $t = 1, \ldots, T$ period.

The key interest is whether nowcasting performance monotonically improves as the t month approaches the updated date of the target variable. Where $L(\cdot)$ is the function constructed by different kind error term of different nowcast error, for example the squared differences, absolute differences and absolute percentage errors corresponding to RMSE, MAE and MAPE. We aimed at knowing whether the nowcast error loss at a point $i+k$ is lower than some earlier point i. Due to the limited space in the study, we used the form of a squared difference similar to the calculation of RMSE when constructing the $L(\cdot)$ function. The null hypothesis is formed of $S(S-1)/2$ moment inequalities for each pairwise comparison of nowcasting points $i+k$ and i:

$$H_0: E[\Delta L_t(\hat{y}_{i+k,t}, \hat{y}_{i,t})] = E\left[L(y_t - \hat{y}_{i+k,t}) - L(y_t - \hat{y}_{i,t})\right] \le 0; \forall i,k (i=1,\ldots,S; k=1,\ldots,S-1), \quad (12)$$

$$H_1: E[(\Delta L_t(\hat{y}_{i+k,t}, \hat{y}_{i,t}))] = E\left[L(y_t - \hat{y}_{i+k,t}) - L(y_t - \hat{y}_{i,t})\right] > 0; \exists i,k (i=1,\ldots,S; k=1,\ldots,S-1), \quad (13)$$

where $\Delta L_t(\hat{y}_{i+k,t}, \hat{y}_{i,t}) = L(y_t - \hat{y}_{i+k,t}) - L(y_t - \hat{y}_{i,t})$ is the difference between the value based on the nowcasting error at the $i+k$ and i, also known as the second moment generated by the $i+k$ and i. The above null hypothesis H_0 means that when the $S(S-1)/2$ moment inequality is tested, as long as there is at least one point later in the nowcasting period that has a larger loss than an earlier horizon, it will be violated.

This study considers all possible $S(S-1)/2$ pairwise moment inequalities in the test, rather than just adjacent inequalities, to detect any violation of H_0. Next, the recursive window is used to generate nowcasting results. Suppose the length of the full sample is T, and we split the length of the in-sample is R, and the out-of-sample is (T-R). The in-sample $\{y_t\}_{t=1}^{R}$ is used for estimation, and the out-of-sample $\{y_t\}_{t=R+1}^{T}$ is used for testing the model's nowcasting performance. We can calculate the following statistics under the null hypothesis H_0:

$$\frac{1}{T-R}\sum_{t=R+1}^{T}\left[L(y_t - \hat{y}_{i+k,t}) - L(y_t - \hat{y}_{i,t})\right] = \frac{1}{T-R}\Delta L_t(\hat{y}_{i+k,t}, \hat{y}_{i,t}), i=1,\ldots,S; k=1,\ldots,S-1. \quad (14)$$

To test the null hypothesis in Equation (12), we use a statistic based on the empirical moment inequalities introduced in Equation (14). That is, the test statistic is the max statistic of the following form:

$$U^* = \max_{i,k\in C_s}\sum_{t=R+1}^{T}\frac{\left[L(y_t - \hat{y}_{i+k,t}) - L(y_t - \hat{y}_{i,t})\right]}{\sqrt{T-R}} = \max_{i,k\in C_s}\sum_{t=R+1}^{T}\frac{\left[\Delta L_t(\hat{y}_{i+k,t}, \hat{y}_{i,t})\right]}{\sqrt{T-R}}, \quad (15)$$

where *k* denotes the cardinality of C_s, $C_s = \{i,k : i=1,...,S, k=1,...,S-i\}$, assuming that $c(\alpha)$ is the corresponding critical value at level *α* from the block multiplier bootstrap (BMB) procedure described in Fosten and Gutknecht (2020). For any *i* and *i*+*k*, the variance of $L(y_t - y_{i+k,t}(\theta_{i+k})) - L(y_t - y_{i,t}(\theta_i))$, $v^2_{i+k,i} \equiv Var(L(y_t - y_{i+k,t}(\theta_{i+k})) - L(y_t - y_{i,t}(\theta_i)))$, should be bounded away from zero to avoid the perfectly correlation between the adjacent horizons' factors when the nowcast times *S* tend to infinity. The above problem is addressed by setting k equal to some deterministic sequence k_s^c to reduce a subset of moment inequalities which are close. Specifically, for some $S > 1$, $k_s^c = \min\left\{k \in \{1,...,S-i\} : v^2_{i+k,i} > c > 0, \forall i = 1,...,S-1\right\}$. k_s^c is the smallest integer *k* in which the variance $v^2_{i+k,i}$ is larger than some constant *c* of any *i* and *i*+*k*.

$k_s^c = 1$ means that the moment inequality of the minimum interval to be tested is $[L(y_t - \hat{y}_{3,t}) - L(y_t - \hat{y}_{1,t})]$, and $k_s^c = 5$ means that the moment inequality of the minimum interval to be tested is $[L(y_t - \hat{y}_{7,t}) - L(y_t - \hat{y}_{1,t})]$. Therefore, when a difference k_s^c is selected, the number of corresponding moment inequality comparison sets C_s will also change.

For the monotonicity test:

$U^* \le c(\alpha) \Rightarrow$ Accept the null hypothesis H_0,there is monotonicity;
$U^* > c(\alpha) \Rightarrow$ Reject the null hypothesis H_0, there is no monotonicity;

$U^* > c(\alpha)$ represents the null hypothesis is rejected; that is, there is at least one point later in the nowcasting period that has a larger loss than an earlier horizon. The critical value $c(\alpha)$ is determined based on the BMB method under the null hypothesis to test the finite sample approximation of the asymptotic distribution of the statistic. Due to space limitations, the specific steps of the BMB algorithm can be found in Fosten and Gutknecht (2020). It should be noted that because the hypothesis of the monotonicity test method is very strict, the existence of monotonicity is considered to be violated if at least one point later in the nowcasting period has a larger loss than an earlier horizon. Therefore, according to Fosten and Gutknecht (2020), when the p-value obtained from the test result is greater than 70%, there is obvious evidence that the nowcasting performance has monotonicity.

The data

Tourism demand is represented by monthly visitor arrivals (VA) from the mainland of China to Hong Kong, Macau, and Taiwan from January 2011 to December 2019, obtained from the Wind Database (http://www.wind.com.cn/). The tourism demand data after December 2019 had greater volatility due to COVID-19 (Wang et al., 2021; Marques et al., 2022). Therefore, the data after December 2019 were not included in this study. Figure 2 shows visitor arrivals from the mainland of China to the three other Greater China destinations (Hong Kong, Macau, Taiwan) over this period.

[Insert Figure 2 about here]

The dependent variables, the eight monthly macroeconomic variables, and the daily search query data used to construct the Baidu index are collected as the determinants of tourism demand in Hong Kong, Macau, and Taiwan. The monthly macroeconomic variables, which range from January 2011 to December 2019, are collected from the Wind Database (http://www.wind.com.cn/) and the CEInet Statistics Database (https://db.cei.cn/). The daily Baidu index data, which range from January 1, 2011 to December 31, 2019, are collected from the Baidu index database (http://index.baidu.com/).

Macroeconomic variables

Economic theory suggests that the tourism price of the destination, the tourism price of competing destinations, and the income of tourists are the most important factors affecting tourism demand (Gunter and Önder, 2015; Song et al., 2003; Song et al., 2009a). Other macroeconomic variables considered in this study include the *EPU*, *CCI*, *CPI*, *CPDs*, and VA_{lags}. All variables are in logarithmic form. The logarithmic transformation of the variables is to convert a non-linear relationship to a linear one in tourism demand analysis (Wu et al., 2017).

The destination of the own price **(** $P_{j,t}$ **):** The tourism price of the destination relative to that of the source market is expected to have a negative effect on tourism demand in the destination. It is usually measured by the relative *CPI* between the destination and the location of origin, adjusted by the exchange rate (Song et al., 2003). In this study, $P_{j,t}$ is calculated as follows:

$$P_{j,t} = (CPI_{j,t} / EX_{j,t}) / (CPI_{ML,t} / EX_{ML,t}),$$

where $j \in \{HK, MO, TW\}$ [①] and the location of origin refers to the mainland of China.

The substitute price **(** $P_{j,s,t}$ **):** $P_{j,s,t}$ is also an important variable often cited by scholars in studying tourism demand. It refers to the living cost of tourists in alternative destinations, and its calculation is similar to tourism price. According to

[①] In this study, Hong Kong, Macau, and Taiwan are abbreviated as HK, MO, TW in Equations, Tables and Figures for simplicity, according to International Standard Norme Internationale: https://www.iso.org/standard/72482.html.

Blake and Cortes-Jiménez (2007), the impact of substitute destinations is included in modelling the tourism demand in the following ways:

(1) The tourists' living cost variable may be specified as the ratio of the destination value to the original value.

(2) The tourists' cost of a living variable may be specified as destination value relative to a weighted average value calculated for a set of alternative destinations or by specifying a separated weighted average substitute destination cost variable.

The substitute price refers to the tourism price in substitute destinations. It is usually measured by the *CPI* of the substitute destination or the weighted average of the *CPI* of a group of alternative destinations. If this price has a positive influence on tourism demand, this would suggest that the price has a substitution effect, whereas a negative effect would indicate a complementary effect (Blake and Cortes-Jiménez, 2007).

This study takes the market share of the alternative destination (the number of tourists) as the weight. The consumer price index is calculated by weighing the consumer price index of each of the four substitute destinations and making corresponding adjustments to the consumer price index. Considering geographic and cultural characteristics, we choose Thailand, Japan, Taiwan and Macau as alternative destinations to Hong Kong. Then, Thailand, Japan, Hong Kong and Macau are chosen as alternative destinations to Taiwan. Thailand, Japan, Hong Kong and Taiwan are chosen as alternative destinations to Macau. In the case of Hong Kong, the substitute destination price is calculated by:

$$P_{HK,s,t} = \sum_{d=1}^{4} (CPI_{d,t} / EX_{d,t}) / (CPI_{HK,t} / EX_{HK,t}) w_{d,t},$$

where $w_{d,t} = Q_{d,t} \Big/ \left(\sum_{d=1}^{4} Q_{d,t} \right)$, $d = 1, 2, 3, 4$ represents Japan, Thailand, Macau, and

Taiwan. The calculation of $P_{j,s,t}$ for Macau and Taiwan is similar to that for Hong Kong. The only difference is that the own price data are omitted from the calculation.

***Industrial production index* ($IP_{ML,t}$):** The income variable is measured by the real GDP in the origin country/region. Given the absence of monthly data on GDP, $IP_{ML,t}$ is chosen to represent tourist income (Goh et al., 2008). The strong correlation between these two indexes (0.97) suggests that $IP_{ML,t}$ is a reasonable substitute for GDP. Tourist income is expected to have a positive impact on tourism demand.

***Consumer price differentials* ($CPDs_{j,t}$):** The consumer price difference is used to evaluate the difference in consumer prices between the destination Hong Kong/ Macau/Taiwan and the source market of the mainland of China:

$$CPDs_{j,t} = CPI_{j,t} / CPI_{ML,t}.$$

***The economic policy uncertainty* ($EPU_{ML,t}$):** $EPU_{ML,t}$ reflects consumers' expectations. This variable measures economic uncertainty induced by the fiscal and monetary policies in the source markets.

Other determinants of tourism demand include transportation costs (normally measured by oil prices), advertising expenditure, the population in the source market, the source market unemployment rate, and one-off events (Wu et al., 2017). These variables are either unavailable or difficult to measure. As such, they have been excluded from this study.

[Insert Table 2 about here]

Table 2 shows the descriptive statistics of the macroeconomic variables included in the tourism demand nowcasting models. The variables are all in logarithmic form.

The skewness, kurtosis, and *J-B* statistics of the eight variables show that most of the variables are normally distributed.

Tourism demand nowcasting requires knowledge of the specific release dates of the macroeconomic variables to incorporate the latest available information. Since the macroeconomic variables are published with different time lags relative to the daily updated Baidu index data, it is necessary to know the specific release dates of macroeconomic variables, as shown in Table 3. It should be noted that the $EPU_{ML,t}$ data website (http://www.policyuncertainty.com/) does not provide a specific release date for the data. It was observed that the $EPU_{ML,t}$ for May was updated on June 4th. Hence, the release date of the $EPU_{ML,t}$ data for each month is set as the 4th of each following month. The timeline for adding the macroeconomic variables to the nowcasting model for Hong Kong is $EPU_{ML,t} \rightarrow CPI_{ML,t} \rightarrow IP_{ML,t} \rightarrow P_{HK,t} = CPDs_{HK,t}$ $\rightarrow VA_{lags,HK,t} \rightarrow CCI_{ML,t} \rightarrow P_{HK,s,t}$. The timeline for Macau is $EPU_{ML,t} \rightarrow P_{MO,s,t} \rightarrow CPI_{ML,t}$ $\rightarrow IP_{ML,t} \rightarrow P_{MO,t} = CPDs_{MO,t} \rightarrow VA_{lags,MO,t} \rightarrow CCI_{ML,t}$. The timeline for Taiwan is $EPU_{ML,t} \rightarrow P_{TW,t} = CPDs_{TW,t} = P_{TW,s,t} \rightarrow CPI_{ML,t} \rightarrow IP_{ML,t} \rightarrow VA_{lags,TW,t} \rightarrow CCI_{ML,t}$. The addition order of $VA_{lags,t}$ and $CCI_{ML,t}$ could be adjusted according to Table 3.

[Insert Table 3 about here]

The most important factors affecting tourism demand are the destination's price $P_{j,t}$, the substitute's price $P_{j,s,t}$, and the income of consumers $IP_{ML,t}$ (Song and Romilly, 2000). We adopt the above three variables to construct Model 1, shown in Table 1. Then the macroeconomic variables, namely $EPU_{ML,t}$, $CPI_{ML,t}$, $CPDs_{j,t}$,

$VA_{lags,j,t}$, and $CCI_{ML,t}$, were added to Model 1 gradually according to their release time to construct Model 2. Taking the tourism demand nowcast process in Hong Kong as an example, the variables contained in Model 2 with the data release dates are shown in Table 4. Specifically, it can be seen from Table 3 that the first data update in each month is that of $EPU_{ML,t}$, which was updated when the nowcast was made on the 4th day of each month, as shown in the second Equation in Table 4. Then the variable $CPI_{ML,t}$ was updated on the 15th day of each month, as shown in the third Equation in Table 4. The variable $CPDs_{HK,t}$ was updated in the nowcasting model on the 20th day of each month, the variable $VA_{lags,HK,t}$ was updated on the 28th of each month, and $CCI_{ML,t}$ was updated on the 29th of each month.

[Insert Table 4 about here]

Baidu index

The Baidu index is compiled based on Wen et al. (2019) and Wen et al. (2020) using keywords related to six aspects of tourism—i.e., dining, attractions, transportation, tours, shopping, and lodging—as search queries. On the basis of the existing research, this study expands the keywords of Hong Kong and the research object to the Macau and Taiwan areas. The purpose is to make the research conclusions more convincing by researching multiple regions. The selected keywords of the Baidu index are obtained from the Baidu website (https://index.baidu.com/). The specific index construction process is as follows. First, the six keywords are chosen. Second, several initial search queries are specified for each aspect of tourism. Third, strongly correlated search queries are collected from a demand map interface provided by the

Baidu index. Finally, correlation analysis is used to check and filter keywords one by one, according to the availability of each search query. Finally, we obtained 166, 75, and 98 search queries for Hong Kong, Macau, and Taiwan. Using Taiwan as an example, the compilation process is shown in Figure 3.

[insert Figure 3 about here]

Empirical results

In this study, the tourism demand nowcasting results in Models 1 through 4 are measured by the traditional loss functions (i.e., the RMSE, MAE, and MAPE).

[Insert Figure 4 about here]

Figure 4 shows the nowcasting performance of visitor arrivals from the mainland of China to Hong Kong, Macau, and Taiwan. The results are arranged in three rows of three figures each, representing the three loss functions in the three destinations. The figures in the first row are the results for Hong Kong, those in the second row are the results for Macau, and those in the last row are the results for Taiwan. Figure 4's first column is the RMSE results, and the second column is the MAE results. Then, the last column is the results of the MAPE. The *x*-axis represents the results of the 31-day nowcasting horizon, the *y*-axis represents the four models, and the *z*-axis represents the values of the traditional loss functions.

First, in general, the nowcasting accuracy of the four models gradually improves. The nowcasting accuracy of Model 1 is the lowest, and the nowcasting accuracy of Model 4 is the best. The nowcasting error decreases with the macroeconomic factors, and the Baidu index factors are added. In other words, as time goes by, the nowcasting error shows a monotonic decreasing trend.

Second, from a horizontal perspective, the tourism demand nowcasting results are the best in Hong Kong, as expected, because the research method used in this study is more suitable for Hong Kong, with its high-dimensional variables. The results of Taiwan are second to Hong Kong. The tourism demand nowcasting effect for Macau is poor.

Finally, from a longitudinal perspective, the evaluation results show that the changing trends of the three loss functions are roughly the same, as the RMSE, MAE, and MAPE are measured by the error or the square of the error between the actual and predictive values. It can be seen from the nowcasting results of the four models under each loss function that regardless of the region or the loss function, the nowcasting error of Model 1 is the largest. After the factors from different data sources are added, the nowcasting accuracy of Model 4 is substantially improved, as can be seen in the three regions. Relative to Model 1, the addition of the macroeconomic factors improves the nowcasting performance of Model 2, and the addition of the Baidu index factors improves the nowcasting performance of Model 3. When the two types of factors are both added, the nowcasting performance of Model 4 is the strongest of all. This can be seen very clearly in the case of Hong Kong.

In short, the nowcasting performance of the model improves substantially after adding the Baidu index factors and the macroeconomic factors. However, the forecasting performance of the models based on either the macroeconomic factors or the Baidu index factors separately is less clear. As can be seen from the above figure, the nowcasting performance of the four models in Hong Kong gradually improves, but the nowcasting ability in Macau and Taiwan does not show a clear trend as the macroeconomic factors and Baidu index factors are added individually in Model 2 and Model 3. Therefore, the combination of macroeconomic factors and Baidu index factors can improve the nowcasting performance. In other words, the nowcasting accuracy of the model can be substantially improved by using richer data.

Model confidence set

Based on the above analysis, this study refers to the model confidence set (MCS) proposed by Hansen et al. (2011). Compared with other testing methods, the advantages of the MCS are that there is no need to select a benchmark model, multiple groups of models can be compared at the same time, and the best model under a certain confidence level can be obtained with one test.

The MCS is convenient when the number of models is large. The bootstrap implementation is simple to use in practice and avoids the need to estimate a high-dimensional covariance matrix.

This study selects two formulations of the null hypothesis that map naturally onto the test statistics:

$$T_R = \max_{u,v \in M} \frac{| d_{i,uv,t} |}{\sqrt{\operatorname{var}(d_{i,uv,t})}}, \tag{16}$$

$$T_{SQ} = \max_{u,v \in M} \frac{(\bar{d}_{i,uv})^2}{\operatorname{var}(d_{i,uv,t})}, \tag{17}$$

where $\bar{d}_{i,uv} = 1/T \sum_{t=H+1}^{H+T} d_{i,uv,t}$ represents the average value of the relative loss function values of the predicted values of model u and model v.

This study compares the results of tourism demand nowcasting in Hong Kong, Macau, and Taiwan under four models and compares the nowcasting performance of the models in each region. To evaluate the model with the best forecasting performance and obtain a robust evaluation result, the MCS is used to perform a forecasting evaluation and illustrate the pros and cons of all models. As noted by Hansen et al. (2011), the larger *p*-value of a model's confidence set test, the higher the nowcasting accuracy of the corresponding model. The closer the *p*-value is to 1, the better the forecast performance of the model. The forecasting performance of Model 4 is the best among all models generated in this study.

[Insert Table 5 about here]

Table 5 shows the MCS results obtained from 5,000 bootstrap simulations for the four models. The first row in the table represents the four models in Hong Kong, Macau, and Taiwan. The first column represents the two statistics of the MCS. The numbers in the table represent the *p*-values of the MCS. The larger the *p*-value, the easier the null hypothesis is to reject. At the same time, a more obvious *p*-value indicates that the model's forecasting performance is better. In general, the *p*-values of the two test statistics under the three loss functions (RMSE, MAE, and MAPE) in Model 4 are all 1.000 in the three regions. It shows that nowcasting performance is best when the macroeconomic factors and Baidu index factors are included together in the model. When the macroeconomic factors and the Baidu index factors are applied separately, the effect on nowcasting performance is less clear; in some cases, nowcasting performance improves more by adding macroeconomic factors than by adding the Baidu index factors, while in other cases, the inverse is true. The MCS results are consistent with the results under the above three loss functions.

Nowcasting monotonicity tests

The monotonicity test results for the four models and for tourism demand nowcasting from the mainland of China to Hong Kong, Macau, and Taiwan are shown in Table 6. The first row shows the number of moment inequalities corresponding to the different interval values in the table. The second row presents the monotonicity statistics and different significance levels obtained for each moment inequality. The *p*-value represents the rejection rate of the monotonicity test. The larger the *p*-value, the better the results of tourism demand nowcasting, as a higher *p*-value indicates a greater probability of monotonicity when nowcasting tourism demand.

[Insert Table 6 about here]

(1) Overall, the results of the nowcasting monotonicity test based on the four models at different intervals show almost no evidence of accepting the null hypothesis, which means there is no monotonicity. The nowcasting performance of most models has a significance level of over 70%, indicating that the models' nowcasting capabilities gradually improve with the addition of the Baidu index factors and macroeconomic factors, except for the results for Macau and Taiwan in Model 4.

Although the nowcasting accuracy of Model 4 is substantially better than that of the other models as shown in Figure 4, its monotonicity test results in Table 6 show that the probability value does not reach 70%. The reason is that the addition of daily data and monthly data helps to improve Model 4's nowcasting performance. However, the high-frequency data used in this study has large fluctuations when the nowcasting model contains mixed frequency data (monthly macroeconomic variables and daily Baidu indexes), making the nowcast results also fluctuate. In addition, the hypothesis of the monotonicity test method is rigorous. The existence of monotonicity is considered to be violated if at least one point later in the nowcasting period has a larger loss than an earlier horizon. Therefore, although the nowcasting performance of Model 4 is better, its monotonicity test results are relatively inferior compared to those of Models 1 through 3.

(2) From a horizontal perspective, the nowcasting monotonicity results of Hong Kong are better than Macau and Taiwan, which means that the nowcasting performance of Hong Kong gradually improves with the nowcasting period's increase. Except for Macau, the monotonicity test of Model 2 significantly supports the null hypothesis, and the nowcasting performance of the other models all have a significance level greater than 70%. These findings are consistent with the results shown by the loss function in Figure 4.

(3) From a longitudinal perspective, the monotonicity test interval is 5. In other words, when the number of monotonicity moment inequalities to be tested is 465, the monotonicity test results of the four models' nowcasting are the best in different regions. This indicates that it is effective to set the interval value based on the correlation between the extracted factors to avoid adjacent nowcasts.

Conclusion

In this study, using macroeconomic variables and the Baidu index as two different data sources, four competing models are specified to examine nowcasts of tourism demand from the mainland of China to Hong Kong, Macau, and Taiwan.

The three most frequently used macroeconomic variables are used to establish Model 1. Based on Model 1, macroeconomic factors are extracted from the remaining five macroeconomic variables. These factors are added to Model 1 to construct Model

2. The Baidu index factors are extracted from the Baidu index and added to Model 1 to construct Model 3. Last, the macroeconomic factors and Baidu index factors are combined into Model 1 to construct Model 4. Through these four models, tourist arrivals nowcasting from the mainland of China to Hong Kong, Macau, and Taiwan is generated. This study uses traditional loss functions (RMSE, MAE, and MAPE) to evaluate the nowcasting performance of the models. This study further analyses nowcasting accuracy through a novel statistical method, the monotonicity test, to test whether a model's nowcasting performance improves gradually as new information is updated and added from different data sources. The following results are obtained.

First, the monotonicity test of tourism demand nowcasting in Hong Kong, Macau, and Taiwan are found to be more intuitive and convincing than traditional loss functions. Unlike previous studies showing that loss functions show the overall size and do not show the trend of the errors between the nowcast value and actual value. This study uses a monotonicity test to eliminate such shortcomings based on the $L(\cdot)$ functions, thus providing a more objective explanation of nowcasting performance.

Second, from the nowcasting results in different models, Model 1 has the lowest nowcasting performance, while adding the macroeconomic factors and Baidu index factors further improves the performance. Although Models 2 and 3, incorporating macroeconomic factors and Baidu index factors, are superior to Model 1 in their nowcasting performance, there is uncertainty regarding this performance. When the two factors are combined with Model 4 to nowcast tourism demand, however, nowcasting performance is substantially improved, and the nowcasting performance contribution becomes the greatest.

Finally, from the monotonicity test results in different regions, Model 1's probability value of the monotonicity test for different regions is the largest compared with the other three models, which is also consistent with the results in Figure 4. The nowcast error of Model 1 gradually decreases during the forecast period, and its downtrend is substantially better than that of Models 2 through 4. Therefore, the probability of the monotonicity test of Model 1 is greater than that of the other three models. However, there are uncertainties in the monotonicity test results of Model 2 and Model 3 in the three regions. Generally speaking, the monotonicity probability value of Hong Kong and Taiwan is greater than Macau.

In summary, the combination of different data sources in this study substantially improves nowcasting performance. Furthermore, the application of the monotonicity test objectively illustrates that nowcasting performance shows a gradual increase with the continuous addition of new information. It is hoped that the results of this study will inform future nowcasting research on tourism demand in Hong Kong, Macau, and Taiwan.

References

Antolini F and Grassini L (2019) Foreign arrivals nowcasting in Italy with Google Trends data. *Quality and Quantity* 53(5): 2385-2401.

Bańbura M, Giannone D, Modugno M, et al. (2013) Now-casting and the real-time data flow. In: Elliott G, Granger C and Timmermann A (eds) *Handbook of economic forecasting* 2 (Part A), pp.195-237.

Bańbura M and Modugno M (2014) Maximum likelihood estimation of factor models on data sets with arbitrary pattern of missing data. *Journal of Applied Econometrics* 29(1): 133-160.

Bangwayo-Skeete PF and Skeete RW (2015) Can Google data improve the forecasting performance of tourist arrivals? Mixed-data sampling approach. *Tourism Management* 46: 454-464.

Barigozzi M, Hallin M, Soccorsi S, et al. (2021) Time-varying general dynamic factor models and the measurement of financial connectedness. *Journal of Econometrics* 222(1): 324-343.

Blake A and Cortes-Jiménez I (2007) The drivers of tourism demand in the UK. A Report Prepared by Christel De Haan Tourism and Travel Research Institute-University of Nottingham for the Department of Culture, Media and Sport. London: UK Parliament.

Bragoli D and Fosten J (2018) Nowcasting Indian GDP. *Oxford Bulletin of Economics and Statistics* 80(2): 259-282.

Brida JG, Matesanz. GD and Segarra V (2020) On the empirical relationship between tourism and economic growth. *Tourism Management* 81: 104131.

Carriere-Swallow Y and Labbe F (2013) Nowcasting with Google Trends in an emerging market. *Journal of Forecasting* 32(4): 289-298.

Castle JL, Fawcett NWP and Hendry DF (2009) Nowcasting is not just contemporaneous forecasting. *National Institute Economic Review* 210(1): 71-89.

Chatziantoniou I, Degiannakis S, Eeckels B, et al. (2016) Forecasting tourist arrivals using origin country macroeconomics. *Applied Economics* 48(27): 2571-2585.

Chernozhukov V, Chetverikov D and Kato K (2019) Inference on causal and structural parameters using many moment inequalities. *The Review of Economic Studies* 86(5): 1867-1900.

Corradi V and Swanson NR (2014) Testing for structural stability of factor augmented forecasting models. *Journal of Econometrics* 182(1): 100-118.

Dimpfl T and Langen T (2018) How unemployment affects bond prices: A mixed frequency Google nowcasting approach. *Computational Economics* 54(2): 551-573.

Emili S, Gardini A and Foscolo E (2020) High spatial and temporal detail in timely prediction of tourism demand. *International Journal of Tourism Research* 22(4): 451-463.

Feng Y, Li G, Sun X, et al. (2019) Forecasting the number of inbound tourists with Google Trends. *Procedia Computer Science* 162: 628-633.

Forni M, Hallin M, Lippi M, et al. (2000) The generalized dynamic-factor model: Identification and estimation. *Review of Economics and Statistics* 82(4): 540-554.

Fosten J and Gutknecht D (2020) Testing nowcast monotonicity with estimated factors. *Journal of Business & Economic Statistics* 38(1): 107-123.

Giannone D, Reichlin L and Small D (2008) Nowcasting: The real-time informational content of macroeconomic data. *Journal of Monetary Economics* 55(4): 665-676.

Goh C, Law R and Mok HMK (2008) Analyzing and forecasting tourism demand: A rough sets approach. *Journal of Travel Research* 46(3): 327-338.

Gunter U and Önder I (2015) Forecasting international city tourism demand for Paris: Accuracy of uni- and multivariate models employing monthly data. *Tourism Management* 46: 123-135.

Hallin M and Liška R (2007) Determining the number of factors in the general dynamic factor model. *Journal of the American Statistical Association* 102(478): 603-617.

Hansen PR, Lunde A and Nason JM (2011) The model confidence set. *Econometrica* 79(2): 453-497.

Hassani H, Heravi S and Zhigljavsky A (2013) Forecasting UK industrial production with multivariate singular spectrum analysis. *Journal of Forecasting* 32(5): 395-408.

Hirashima A, Jones J, Bonham CS, et al. (2017) Forecasting in a mixed up world: Nowcasting Hawaii tourism. *Annals of Tourism Research* 63: 191-202.

Höpken W, Eberle T, Fuchs M, et al. (2019) Google Trends data for analysing tourists' online search behaviour and improving demand forecasting: The case of Åre, Sweden. *Information Technology and Tourism* 21(1): 45-62.

Hu M, Xiao M and Li H (2021) Which search queries are more powerful in tourism demand forecasting: Searches via mobile device or PC? *International Journal of Contemporary Hospitality Management* 33(6): 2022-2043.

Huang B and Hao H (2021) A novel two-step procedure for tourism demand forecasting. *Current Issues in Tourism* 24(9): 1199-1210.

Jackman M and Naitram S (2015) Research note: Nowcasting tourist arrivals in Barbados - just Google it! *Tourism Economics* 21(6): 1309-1313.

Jiao EX and Chen JL (2019) Tourism forecasting: A review of methodological developments over the last decade. *Tourism Economics* 25(3): 469-492.

Jiao X, Li G and Chen JL (2020) Forecasting international tourism demand: A local spatiotemporal model. *Annals of Tourism Research* 83: 1-12.

Knotek ES, II and Zaman S (2017) Nowcasting US headline and core inflation. *Journal of Money Credit and Banking* 49(5): 931-968.

Kulshrestha A, Krishnaswamy V and Sharma M (2020) Bayesian BILSTM approach for tourism demand forecasting. *Annals of Tourism Research* 83: 1-19.

Li G (2009) Tourism demand modeling and forecasting: A review of literature related to Greater China. *Journal of China Tourism Research* 5(1): 2-40.

Li G, Song H and Witt SF (2005) Recent developments in econometric modeling and forecasting. *Journal of Travel Research* 44(1): 82-99.

Li X, Pan B, Law R, et al. (2017) Forecasting tourism demand with composite search index. *Tourism Management* 59: 57-66.

Li Y, Xu M, Wen X, et al. (2020) The role of internet search index for tourist volume prediction based on GDFM model. *Tehnicki vjesnik - Technical Gazette* 27(2): 576-582.

Marcellino M, Porqueddu M and Venditti F (2016) Short-term GDP forecasting with a mixed-frequency dynamic factor model with stochastic volatility. *Journal of Business & Economic Statistics* 34(1): 118-127.

Marques CP, Guedes AS and Bento R (2022) Tracking changes in tourism demand with point-of-sale data: The case of Portugal. *Tourism and Hospitality Research.* Epub ahead of print Feb 2022. DOI: 10.1177/14673584221075175.

Nor ME, Nurul AIM, Rusiman MS, et al. (2018) A hybrid approach on tourism demand forecasting. In: *International Seminar on Mathematics and Physics in Sciences and Technology*, University of Tun Hussein Onn Malaysia, April, 23, pp.12-34.

Pan B, Wu CD and Song H (2012) Forecasting hotel room demand using search engine data. *Journal of Hospitality & Tourism Technology* 3(3): 196-210.

Pan B and Yang Y (2017) Forecasting destination weekly hotel occupancy with big data. *Journal of Travel Research* 56(7): 957-970.

Pan Y, Xiao Z, Wang X, et al. (2017) A multiple support vector machine approach to stock index forecasting with mixed frequency sampling. *Knowledge-Based Systems* 122: 90-102.

Patton AJ and Timmermann A (2012) Forecast rationality tests based on multi-horizon bounds. *Journal of Business & Economic Statistics* 30(1): 1-17.

Siliverstovs B and Wochner DS (2018) Google Trends and reality: Do the proportions match? Appraising the informational value of online search behavior: Evidence from Swiss tourism regions. *Journal of Economic Behavior & Organization* 145: 1-23.

Silva ES, Ghodsi Z, Ghodsi M, et al. (2017) Cross country relations in European tourist arrivals. *Annals of Tourism Research* 63: 151-168.

Song H, Dwyer L, Li G, et al. (2012) Tourism economics research: A review and assessment. *Annals of Tourism Research* 39(3): 1653-1682.

Song H and Li G (2008) Tourism demand modelling and forecasting - A review of recent research. *Tourism Management* 29(2): 203-220.

Song H, Li G, Witt SF, et al. (2011) Forecasting tourist arrivals using time-varying parameter structural time series models. *International Journal of Forecasting* 27(3): 855-869.

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

Song H and Romilly P (2000) An empirical study of outbound tourism demand in the UK. *Applied Economics* 32(5): 611-624.

Song H, Witt SF and Li G (2009a) The advanced econometrics of tourism demand. *Tourism Management* 31: 447-451.

Song H, Witt SF, Wong KKF, et al. (2009b) An empirical study of forecast combination in tourism. *Journal of Hospitality and Tourism Research* 33(1): 3-29.

Song H, Wong KKF and Chon KKS (2003) Modelling and forecasting the demand for Hong Kong tourism. *International Journal of Hospitality Management* 22(4): 435-451.

UNWTO (2020) International tourism growth continues to outpace the global economy. Available at: http://www.unwto.org/international-tourism-growth-continues-to-outpace-the-economy (accessed July 30, 2021).

Valdivia A and Monge-Corella S (2010) Diseases tracked by using Google Trends, Spain. *Emerging Infectious Diseases* 16(1): 168-168.

Volchek K, Liu A, Song H, et al. (2019) Forecasting tourist arrivals at attractions: Search engine empowered methodologies. *Tourism Economics* 25(3): 425-447.

Wan SK and Song HY (2018) Forecasting turning points in tourism growth. *Annals of Tourism Research* 72: 156-167.

Wang Y, Zhang H, Gao W, et al. (2021) Spillover effects from news to travel and leisure stocks during the COVID-19 pandemic: Evidence from the time and frequency domains. *Tourism Economics.* Epub ahead of print Dec 2021. DOI: 10.1177/13548166211058497.

Wen L, Liu C and Song H (2019) Forecasting tourism demand using search query data: A hybrid modelling approach. *Tourism Economics* 25(3): 309-329.

Wen L, Liu C, Song H, et al. (2020) Forecasting tourism demand with an improved mixed data sampling model. *Journal of Travel Research* 60(2): 336-353.

WTTC (2019) Economic impact reports. Available at: http://wttc.org/Research/Economic-Impact (accessed July 30, 2021).

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

Yang X, Pan B, Evans JA, et al. (2015) Forecasting Chinese tourist volume with search engine data. *Tourism Management* 46: 386-397.

Table 1. Specification of the nowcasting models of tourism demand.

Model	Equation
Model 1	$\ln VA^M_{j,t} = \beta_0 + \beta^M_p \ln P^M_{j,t-1} + \beta^M_{ip} \ln IP^M_{ML,t-1} + \beta^M_{p_s} \ln P^M_{j,s,t-1} + u_{j,t}$
Model 2	$\ln VA^M_{j,t} = \beta_0 + \beta^M_p \ln P^M_{j,t-1} + \beta^M_{ip} \ln IP^M_{ML,t-1} + \beta^M_{p_s} \ln P^M_{j,s,t-1} + \sum_{i=1}^{n_M} \beta^M_i \ln x^M_{i,t-1} + u_{j,t}$
Model 3	$\ln VA^M_{j,t} = \beta_0 + \beta^M_p \ln P^M_{j,t-1} + \beta^M_{ip} \ln IP^M_{ML,t-1} + \beta^M_{p_s} \ln P^M_{j,s,t-1} + \sum_{i=1}^{n_D} \beta^D_i W^D_i (L^{1/m}, \theta_i) \mathrm{X}^D_{i,t} + u_{j,t}$
Model 4	$\ln VA^M_{j,t} = \beta_0 + \beta^M_p \ln P^M_{j,t-1} + \beta^M_{ip} \ln IP^M_{ML,t-1} + \beta^M_{p_s} \ln P^M_{j,s,t-1}$ $+ \sum_{i=1}^{n_M} \beta^M_i \ln x^M_{i,t-1} + \sum_{i=1}^{n_D} \beta^D_i W^D_i (L^{1/m}, \theta_i) \mathrm{X}^D_{i,t} + u_{j,t}$

Note: $VA^M_{j,t}$ refers to monthly visitor arrivals from the mainland of China to destination j at time t; $x^M_{i,t-1}$ is the i^{th} monthly macroeconomic variable ($EPU^M_{ML,t-1}, CPI^M_{ML,t-1}, CPDs^M_{j,t-1}, VA^M_{lags,j,t-1}, CCI^M_{ML,t-1}$) at time t-1; $i = 1,...,n_M$; $\mathrm{X}^D_{i,t}$ is the i^{th} daily Baidu index factor; $i = 1,...,n_D$, and $j \in \{HK, MO, TW\}$. Moreover, the macroeconomic variables are logarithmically processed, and the superscripts in the formula represent the frequency of the data; i.e., superscripts M and D mean the variables are sampled in monthly and daily frequencies, respectively.

Table 2. Descriptive statistics.

Variables		*Mean*	*SD*	*Min*	*Max*	*Skewness*	*Kurtosis*	*J-B*
$EPU_{ML, t}$		5.465	0.780	3.264	6.878	-0.215	2.729	1.165
$CCI_{ML, t}$		4.690	0.079	4.575	4.841	0.603	2.008	10.966***
$IP_{ML, t}$		5.228	0.212	4.628	5.638	-0.349	2.500	3.315
$CPI_{ML, t}$		4.630	0.012	4.613	4.668	1.499	4.695	53.362***
$CPDs_{j, t}$	$CPDs_{HK, t}$	-0.037	0.087	-0.225	0.075	-0.656	2.259	10.222***
	$CPDs_{MO, t}$	-0.133	0.108	-0.368	-0.003	-0.735	2.302	11.926***
	$CPDs_{TW, t}$	-0.033	0.032	-0.114	0.009	-1.056	3.317	20.527***
$P_{j, t}$	$P_{HK, t}$	1.146	0.085	0.921	1.288	-0.997	3.762	20.502***
	$P_{MO, t}$	0.078	0.099	-0.182	0.226	-1.132	3.729	25.472***
	$P_{TW, t}$	1.516	0.059	1.353	1.645	-0.840	4.269	19.950***
$P_{j, s, t}$	$P_{HK, s, t}$	7.794	0.363	7.114	8.235	-0.436	1.596	12.286***
	$P_{MO, s, t}$	7.518	0.348	6.924	8.206	-0.120	1.624	8.781**
	$P_{TW, s, t}$	1.516	0.059	1.353	1.645	-0.840	4.269	19.950***
$VA_{logs, j, t}$	$VA_{lags, HK, t}$	15.045	0.232	14.470	15.528	-0.606	2.889	6.676**
	$VA_{lags, MO, t}$	14.358	0.191	13.997	14.789	0.269	2.536	2.271
	$VA_{lags, TW, t}$	12.382	0.332	11.482	12.915	-0.731	3.223	9.836***

Note: $EPU_{ML, t}$, $CCI_{ML, t}$, $IP_{ML, t}$, and $CPI_{ML,t}$ respectively represent economic policy uncertainty in the mainland of China, *ML* means the mainland of China, consumer confidence index, industrial production index, and consumer price index; $CPDs_{j, t}$,$P_{j, t}$, $P_{j, s, t}$, and $VA_{lags, j, t}$ respectively represent the consumption difference index, destination price, competitive alternative price, and the lagged form of tourist arrivals in the three other Greater China destinations, $j \in \{HK, MO, TW\}$.

Table 3. Date of the first release of data regarding the three other Greater China destinations in 2019M1-2019M12.

Month		*Jan.*	*Feb.*	*Mar.*	*Apr.*	*May.*	*Jun.*	*Jul.*	*Aug.*	*Sept.*	*Oct.*	*Nov.*	*Dec.*
$EPU_{ML,t}$		19/02/04	19/03/04	19/04/04	19/05/04	19/06/04	19/07/04	19/08/04	19/09/04	19/10/04	19/11/04	19/12/04	20/01/04
$CCI_{ML,t}$		19/03/11	19/03/14	19/04/30	19/05/29	19/07/02	19/07/31	19/09/02	19/09/29	19/10/31	19/12/03	19/12/31	20/02/07
$IP_{ML,t}$		19/02/17	19/03/15	19/04/17	19/05/15	19/06/14	19/07/15	19/08/15	19/09/16	19/10/18	19/11/14	19/12/16	20/01/21
$CPI_{ML,t}$		19/02/15	19/03/09	19/04/11	19/05/09	19/06/12	19/07/10	19/08/09	19/09/10	19/10/16	19/11/09	19/12/09	20/01/10
$CPDs_{j,t}$	$CPDs_{HK,t}$	19/02/20	19/03/23	19/04/23	19/05/21	19/06/21	19/07/22	19/08/20	19/09/20	19/10/22	19/11/21	19/12/20	20/01/21
	$CPDs_{MO,t}$	19/02/21	19/03/20	19/04/22	19/05/21	19/06/12	19/07/17	19/08/19	19/09/20	19/10/21	19/11/25	19/12/19	20/01/21
	$CPDs_{TW,t}$	19/03/08	19/05/07	19/05/07	19/05/07	19/07/05	19/08/06	19/09/06	19/10/05	19/12/05	20/02/12	20/02/12	20/02/12
$P_{j,t}$	$P_{HK,t}$	19/02/20	19/03/23	19/04/23	19/05/21	19/06/21	19/07/22	19/08/20	19/09/20	19/10/22	19/11/21	19/12/20	20/01/21
	$P_{MO,t}$	19/02/21	19/03/20	19/04/22	19/05/21	19/06/12	19/07/17	19/08/19	19/09/20	19/10/21	19/11/25	19/12/19	20/01/21
	$P_{TW,t}$	19/03/08	19/05/07	19/05/07	19/05/07	19/07/05	19/08/06	19/09/06	19/10/05	19/12/05	20/02/12	20/02/12	20/02/12
$P_{j,s,t}$	$P_{HK,s,t}$	19/03/08	19/05/07	19/05/07	19/05/07	19/07/05	19/08/06	19/09/06	19/10/05	19/12/05	20/02/12	20/02/12	20/02/12
	$P_{MO,s,t}$	19/03/08	19/05/07	19/05/07	19/05/07	19/07/05	19/08/06	19/09/06	19/10/05	19/12/05	20/02/12	20/02/12	20/02/12
	$P_{TW,s,t}$	19/03/08	19/05/07	19/05/07	19/05/07	19/07/05	19/08/06	19/09/06	19/10/05	19/12/05	20/02/12	20/02/12	20/02/12
$VA_{lags,j,t}$	$VA_{lags,HK,t}$	19/03/01	19/04/01	19/04/29	19/05/31	19/06/28	19/07/31	19/08/30	19/09/28	19/10/31	19/11/30	19/12/31	20/01/31
	$VA_{lags,MO,t}$	19/02/26	19/03/22	19/04/23	19/05/23	19/06/21	19/07/22	19/08/21	19/09/24	19/10/23	19/11/23	19/12/21	20/01/21
	$VA_{lags,TW,t}$	19/03/11	19/03/27	19/04/30	19/05/30	19/06/25	19/07/25	19/08/26	19/09/25	19/10/25	19/11/25	19/12/30	20/02/27

Note: $EPU_{ML,t}$, $CCI_{ML,t}$, $IP_{ML,t}$, and $CPI_{ML,t}$ respectively represent economic policy uncertainty in the mainland of China, *ML* means the mainland of China, consumer confidence index, industrial production index, and consumer price index; $CPDs_{j,t}$, $P_{j,t}$, $P_{j,s,t}$, and $VA_{lags,j,t}$ respectively represent the consumption difference index, destination price, competitive alternative price, and the lagged form of tourist arrivals in the three other Greater China destinations.

Table 4. The changing nowcasting equation in Model 2 with data updated in Hong Kong

Update time in each month	Equation
1st~3rd day	$\ln VA^M_{HK,t} = \beta_0 + \beta^M_p \ln P^M_{HK,t-1} + \beta^M_{ip} \ln IP^M_{ML,t-1} + \beta^M_{p_s} P^M_{HK,s,t-1} + \beta^M_{epu} \ln EPU^M_{ML,t-2} + \beta^M_{cpi} \ln CPI^M_{ML,t-2} + \beta^M_{cpds} \ln CPDs^M_{HK,t-2} + \beta^M_{va} \ln VA^M_{lags,HK,t-2} + \beta^M_{cci} \ln CCI^M_{ML,t-2} + u_{HK,t}$
4th~14th day	$\ln VA^M_{HK,t} = \beta_0 + \beta^M_p \ln P^M_{HK,t-1} + \beta^M_{ip} \ln IP^M_{ML,t-1} + \beta^M_{p_s} P^M_{HK,s,t-1} + \beta^M_{epu} \underline{\ln EPU^M_{ML,t-1}} + \beta^M_{cpi} \ln CPI^M_{ML,t-2} + \beta^M_{cpds} \ln CPDs^M_{HK,t-2} + \beta^M_{va} \ln VA^M_{lags,HK,t-2} + \beta^M_{cci} \ln CCI^M_{ML,t-2} + u_{HK,t}$
15th~19th day	$\ln VA^M_{HK,t} = \beta_0 + \beta^M_p \ln P^M_{HK,t-1} + \beta^M_{ip} \ln IP^M_{ML,t-1} + \beta^M_{p_s} P^M_{HK,s,t-1} + \beta^M_{epu} \ln EPU^M_{ML,t-1} + \beta^M_{cpi} \underline{\ln CPI^M_{ML,t-1}} + \beta^M_{cpds} \ln CPDs^M_{HK,t-2} + \beta^M_{va} \ln VA^M_{lags,HK,t-2} + \beta^M_{cci} \ln CCI^M_{ML,t-2} + u_{HK,t}$
20th~27th day	$\ln VA^M_{HK,t} = \beta_0 + \beta^M_p \ln P^M_{HK,t-1} + \beta^M_{ip} \ln IP^M_{ML,t-1} + \beta^M_{p_s} P^M_{HK,s,t-1} + \beta^M_{epu} \ln EPU^M_{ML,t-1} + \beta^M_{cpi} \ln CPI^M_{ML,t-1} + \beta^M_{cpds} \underline{\ln CPDs^M_{HK,t-1}} + \beta^M_{va} \ln VA^M_{lags,HK,t-2} + \beta^M_{cci} \ln CCI^M_{ML,t-2} + u_{HK,t}$
28th~31st day	$\ln VA^M_{HK,t} = \beta_0 + \beta^M_p \ln P^M_{HK,t-1} + \beta^M_{ip} \ln IP^M_{ML,t-1} + \beta^M_{p_s} P^M_{HK,s,t-1} + \beta^M_{epu} \ln EPU^M_{ML,t-1} + \beta^M_{cpi} \ln CPI^M_{ML,t-1} + \beta^M_{cpds} \ln CPDs^M_{HK,t-1} + \beta^M_{va} \underline{\ln VA^M_{lags,HK,t-1}} (or \ln VA^M_{lags,HK,t-2}) + \beta^M_{cci} \underline{\ln CCI^M_{ML,t-1}} (or \ln CCI^M_{ML,t-2}) + u_{HK,t}$

Notes: Since the release date of the macroeconomic variables in each month is different, the specific number of periods in which variables can be added to the formula is determined according to the specific release dates of the available variables, shown in Table 3. The underlined part of the formula indicates the newly updated data that has been added in the current period. The addition order of $VA_{lags,HK,t}$ and $CCI_{ML,t}$ could be adjusted according to Table 3.

Table 5. Model confidence set test results.

Loss \ Model		*Hong Kong* 1	2	3	4	*Taiwan* 1	2	3	4	*Macau* 1	2	3	4
RMSE	T_{SQ}	0.635	0.000	0.277	1.000	0.000	0.527	0.000	1.000	0.000	0.855	0.000	1.000
	T_R	0.000	0.000	0.000	1.000	0.000	0.534	0.000	1.000	0.000	0.855	0.000	1.000
MAE or MAPE	T_{SQ}	0.000	0.000	0.000	1.000	0.000	0.000	0.000	1.000	0.000	0.488	0.000	1.000
	T_R	0.000	0.000	0.000	1.000	0.000	0.000	0.000	1.000	0.000	0.478	0.000	1.000

Notes: The closer the p-value is to 1, the more prominent the model's tourism demand nowcasting performance. 1, 2, 3, and 4 in the above table represent Model 1, Model 2, Model 3, and Model 4, respectively.

Table 6. Monotonicity test results of tourism demand nowcasting in the three other Greater China destinations.

Model		465				435				30			
		U^*	**90%**	**95%**	p	U^*	**90%**	**95%**	p	U^*	**90%**	**95%**	p
Model 1	***HK***	0.000	0.996	1.315	**0.927**	0.000	1.135	1.493	**0.917**	0.000	0.680	0.832	**0.900**
	MO	0.406	1.761	2.082	**0.719**	0.406	1.705	2.006	0.679	0.391	1.536	1.906	0.684
	TW	0.000	1.237	1.453	**0.855**	0.000	1.212	1.355	**0.857**	0.000	1.150	1.395	**0.825**
Model 2	***HK***	1.097	3.441	4.181	**0.767**	1.097	3.408	4.095	**0.707**	1.013	3.395	4.144	**0.825**
	MO	2.655	3.584	4.283	0.236	2.655	4.440	5.061	0.321	1.588	2.387	2.773	0.358
	TW	0.365	2.625	3.065	**0.960**	0.365	2.454	3.121	**0.970**	0.365	2.092	2.451	**0.953**
Model 3	***HK***	1.007	9.804	10.901	**0.965**	1.007	9.454	10.88	**0.942**	0.796	7.071	8.511	**0.782**
	MO	3.140	9.961	11.30	**0.769**	3.140	9.698	11.02	**0.782**	2.120	7.411	8.309	**0.719**
	TW	0.981	8.555	9.802	**0.877**	0.981	7.994	9.455	**0.880**	0.675	7.202	8.653	**0.812**
Model 4	***HK***	3.317	10.80	13.12	**0.712**	3.317	10.39	12.77	**0.827**	2.459	7.630	9.072	0.612
	MO	3.103	9.978	11.69	0.674	3.103	9.914	11.40	0.694	2.153	7.778	9.584	0.454
	TW	3.602	8.362	9.999	0.624	3.602	8.587	10.48	**0.724**	2.860	6.005	7.032	0.587

Notes: “Model 1” represents the traditional three-variable model. “Model 2” represents the model based on Model 1 constructed by extracting macroeconomic factors from five macroeconomic variables. “Model 3” represents the model built based on Model 1 by adding the Baidu index factors. “Model 4” represents Model 1, with both macroeconomic factors and Baidu index factors.

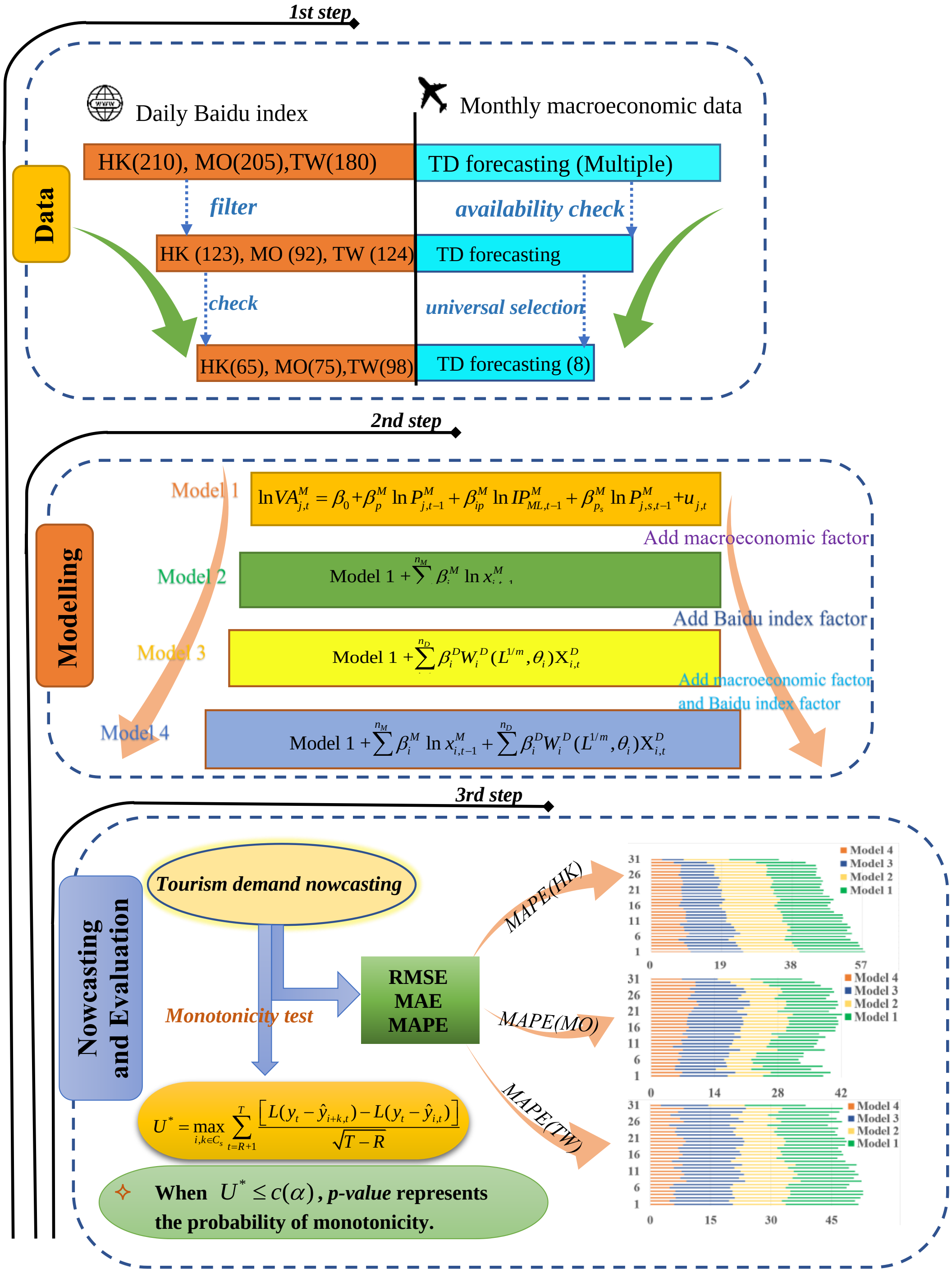

Figure 1. The research framework.

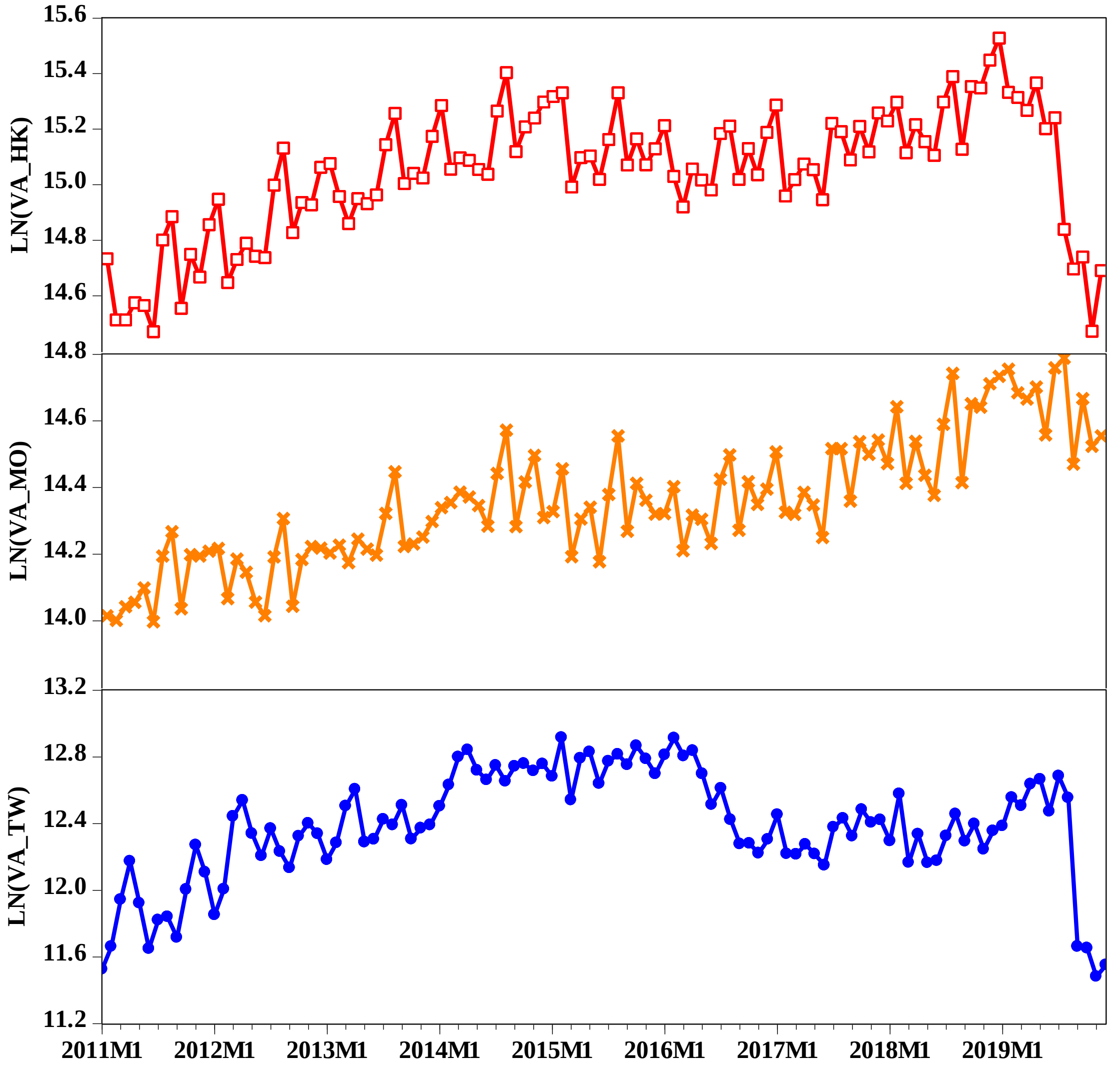

Figure 2. Visitor arrivals from the mainland of China to Hong Kong, Macau, and Taiwan.

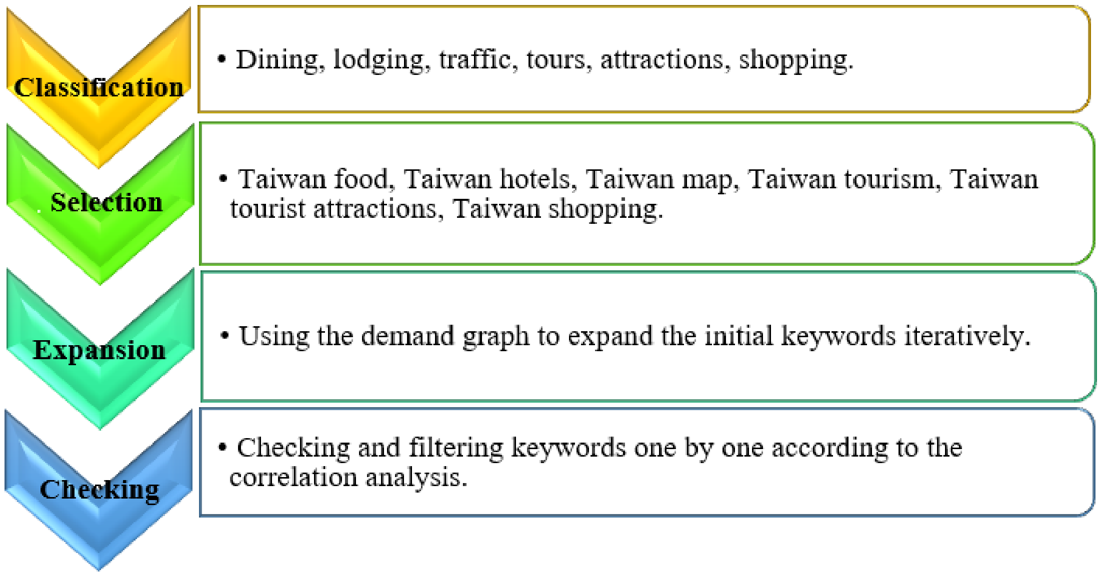

Figure 3. Keyword selection process (the case of Taiwan).

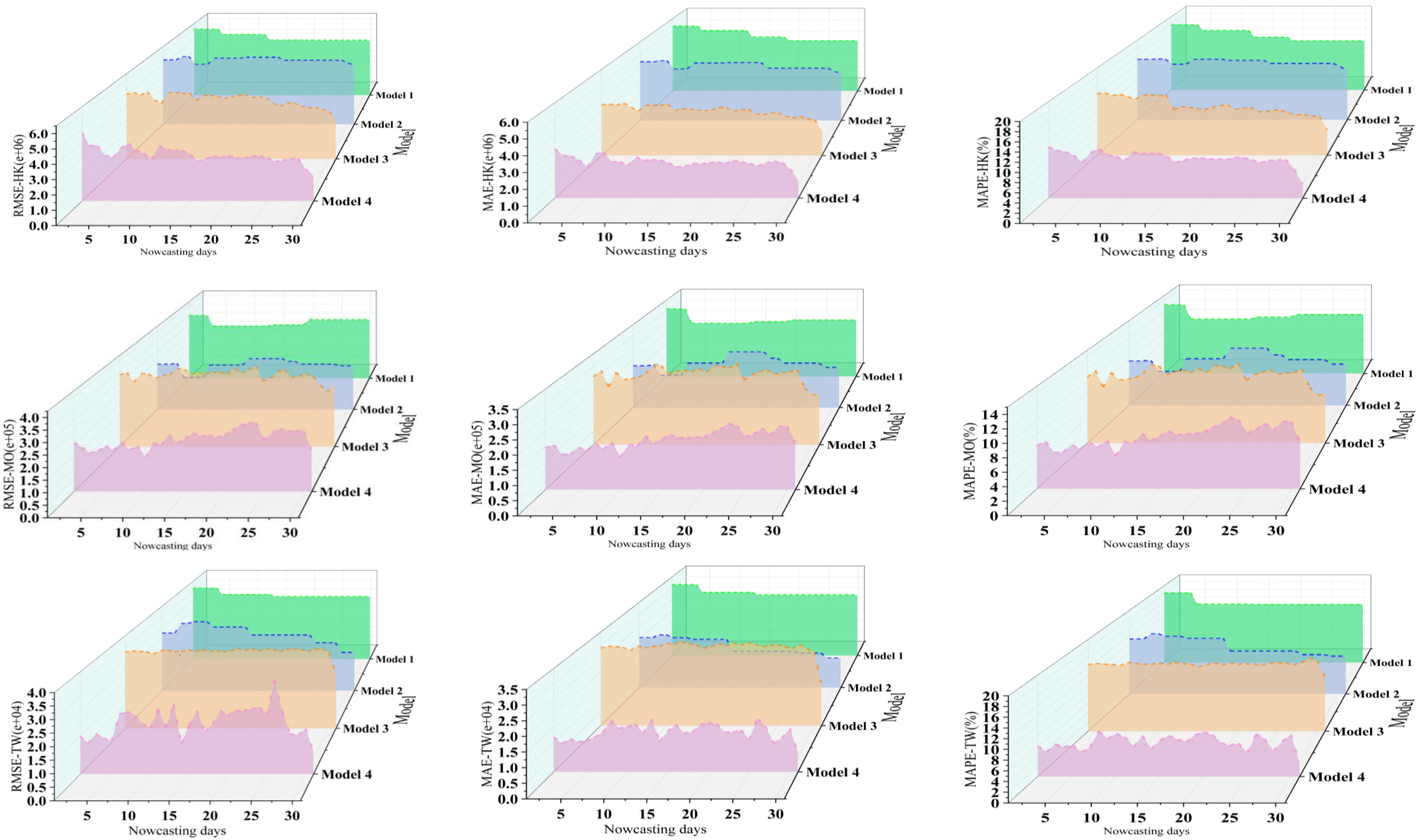

Figure 4. Nowcasting evaluation under different models.