

**RESEARCH ARTICLE**

WILEY

High spatial and temporal detail in timely prediction of tourism demand

Silvia Emili¹ | Attilio Gardini¹ | Enrico Foscolo²¹Department of Statistical Sciences, University of Bologna, Rimini, Italy²School of Economics and Management, Free University of Bozen, Bolzano, Italy**Correspondence**

Silvia Emili, Department of Statistical Sciences, University of Bologna, Piazzetta Teatini, 10, Rimini 47921, Italy.

Email: silvia.emili2@unibo.it**Abstract**

What happens in forecasting problems when high frequency and high spatial detail data encounter significant publication delays? In this paper, we consider a monthly dynamic panel data model, augmented by Google Trends search query volume data, for tourism demand forecasting at high spatial detail, in which one of the main aspects is represented by a publication delay ranging from 8 to 15 months. Some findings in the tourism literature already specify forecasting/nowcasting applications considering a realistic time delay but not for more than 3 months.

KEYWORDS

dynamic panel data model, Google Trends, high spatial detail, weak tourist' flows

1 | INTRODUCTION

Over the past decades, the understanding of tourism dynamics and trend forecasting has emerged as fundamental to enhancing management planning and policy strategies, especially at mass or multiproduct tourism destinations (Frechtling, 2001). In this context, two crucial aspects of the published data are the temporal and the spatial details. Unfortunately, official information on tourism activity is usually published with long delays and as aggregated (areal) measures, undermining the efficiency of the decision-making processes (Gunter & Onder, 2015). Although different statistical and econometric instruments and solutions have been widely investigated, in the field of tourism, several aspects regarding the availability of temporal and spatial detailed data have been almost completely ignored. This paper aims to fill these gaps and create a realistic prediction framework by evaluating a dynamic panel data model augmented by data that are well suited from a temporal and spatial perspective to forecasting and nowcasting the tourism demand at a fine territorial detail (i.e., at the municipality level).

Panel data analysis has not been widely employed in tourism research (Garín-Muñoz, 2004; Narajan et al., 2010; Sequeira & Maçãs-Nunes, 2008), especially with variables observed at frequencies higher than yearly. The availability of informative data is crucial in these cases. Recently, in order to overcome the shortage of official information, the literature has focused on augmenting the model directly from Big Data (for a review, see Hassani & Silva, 2015). In

fact, from a business and economic point of view, potential customers, who are online, produce a large amount of cost-free, informative, timely, and spatially detailed market information, which can be analysed in real time (*ex ante*). Nowadays, the predictive power of web-related data has generated a strong interest in modelling by means of augmented forecasting time-series specifications with a wide set of variables on Internet volumes. This is an important support for innovation, competition, and productivity (Manyika et al., 2011). Tourism research has seized this opportunity to investigate both tourism demand (analysis of a tourist's desires and preferences, e.g., Song & Liu, 2017; Chen et al., 2019) and supply (investigating knowledge of competitors' interests and attractiveness, market position, and information about target product design, e.g., De Mauro, Greco, & Karuga, 2015).

The Google Trends index (GT index) for search query data has received great attention as a result of its ability to predict upcoming trends in different economic sectors (Askatas & Zimmermann, 2009; D'Amuri & Marcucci, 2009; Ginsberg et al., 2009), as well as in the tourism market. The increasing number of tourists exploiting online channels for information on holidays and purchases guarantees that the GT index is a representative proxy for *ex ante* tourism demand of a destination (Choi & Varian, 2009); it also generates a significant strand of tourism literature employing search query data, starting from basic specification (e.g., Varian, 2014; Rivera, 2016) to advanced instruments such as mixed-data sampling equations (Bangwayo-Skeete & Skeete, 2015), linear dynamic models (Rivera, 2016), and

generalized dynamic factor models (Li, Pan, Law, & Huang, 2017). Most of these studies, however, focus primarily on macro tourism areas (i.e., nations or metropolitan areas), without considering data realization and real-time vintages. Camacho and Pacce (2017) are the only exception to the tourism literature with their consideration of publication delays, data releases, and updates. The authors focus on forecasting the demand for Spain at a national level, dealing with a 3-month publication delay of official data. However, enhancing the territorial detail of the information causes a clear increase in this delay.

To overcome these shortcomings in forecasting tourism demand, this study incorporates the following novelties.

1.1 | Fine territorial detail

The first aspect to be considered and enhanced, in this paper, is the prediction of fine territorial detailed data (i.e., municipalities). In the literature, analysis of a specific small area is rare (Goffi & Cucculelli, 2019), and the finest spatial detail considered by scholars is metropolitan cities or cities characterized by high tourism flows (Gunter & Onder, 2015, 2016; Emili, Figini, & Guizzardi, 2019). Small towns, where the tourism industry is at the core of the local economy, may benefit from an accurate analysis of the dynamics of tourism flows to a larger extent than metropolitan areas. This issue is highly relevant to both (a) seasonal destinations, where the predictability of tourism arrivals is fundamental to destination management; and (b) areas characterized by different attractions (i.e., clusters of municipalities) responding to alternative segments of the tourism market.

1.2 | Panel data model

A second important aspect in this paper is found in the methodological approach. In this sense, seasonality, multiproduct destinations, and spatial proximity require coordinated marketing and management actions to deal with and control for potential tourism pull factors in a timely manner. To account for these issues, we suggest using an econometric panel approach, where municipalities within an area represent the unit of the analysis. Song and Li (2008) highlight panel data analysis of tourism demand as an approach embodying “some advantages over the time series econometric models. It incorporates much richer information from both time series and cross sectional data. This approach also reduces the problem of multicollinearity and provides more degrees of freedom in the model estimation”.

1.3 | Nowcasting approach with real (long) publication delay and GT data

Encouraged by the temporal mismatch between timely prediction for hospitality-firm management and tourism-destination urban policies (Jackman & Nairam, 2015) and long delays in the publication of

official statistics, the third contribution of this paper is the forecasting/nowcasting schemes with horizons ranging from 8 to 15 months. This is in line with real availability of fine territorial tourism data. Introducing ex ante demand variable information, on the basis of selected keywords of a Google search-volume index, we estimate timely measures of production activities at the destination level with the inclusion of GT data. In addition, the panel nowcasting model allows us to account for the possible spatial relationships among the tourism destinations in the area, thereby improving the quality of the predictions.

In Italy, foreign visitors bring in €37.2 billion (7.5% of total Italian exports), supporting 5.5% of the direct and 12.6% of total employment. The heterogeneous geographical wealth of Italy along with a rich offering of tourism attractions leads to tourism earnings playing an important role in both the public and private budgets of the local areas. The province of Rimini can be considered one of the most representative examples in Italy of local territories strongly connected to tourism earnings.

The province of Rimini is an interesting setting to evaluate the proposed panel specification for several reasons. First, it is a top Italian tourism destination—ranking second for tourism GDP per capita—where 30% of the total product in the whole area is provided by the tourism sector.¹ Second, it comprises different typologies of municipalities, from the cities located on the Adriatic Sea coast (from north to south, Bellaria, Rimini, Riccione, Misano, and Cattolica) to the inland towns, rich in historical attractions, boasting Romans relics and Mediaeval castles. Furthermore, Rimini area is a mass and multiproduct destination offering solutions to different demand segments including leisure, meeting and conference, sport, and culture.

The remainder of this paper is organized as follows. Section 2 reviews the forecasting and nowcasting literature, especially focusing on the use of search query volumes to enhance accuracy and a lack of interest in publication delays and data vintages. In Section 3, we specify the panel data forecasting model, and we present the case study. Section 4 reports results of both forecasting and nowcasting applications. Finally, in Section 5, we analyse the main findings and the related managerial implications.

2 | LITERATURE REVIEW

2.1 | Tourism forecasting and big data

In general, the tourism forecasting literature is composed of two strands (Song & Li, 2008): econometric models (Uysal & O’Leary, 1986; Lim & McAleer, 2001; Rossellò-Nadal, 2001; Song et al., 2009) and pure time-series models (Witt & Witt, 1991; Hyndman et al., 2002; Gil-Alana, 2005; Athanasopoulos et al., 2011). In this strand of the literature, the first possible causes of failure in tourism modelling and forecasting might be the high spatial substitutability among neighbouring cities or destinations with comparable tourism attractions. A second source of failure in forecasting time-series tourism data may be found in the high unpredictability of tourism choices, which are contingent on war, terrorism, and geopolitical instability in addition to

atmospheric conditions, environmental crises, fad swings, and so forth. To address these possible causes of failure in tourism modelling and forecasting, we combine the exploration of (a) the use of Big Data generated by Internet traffic in order to assess changes in tourism flows from those factors a timely manner and (b) the inclusion of these data in the context of panel data models, accounting for the volatility due to spillovers from substitutable tourism destinations and innovations resulting from changes in tourist preferences.

As clearly pointed out by Song and Liu (2017), in the framework of tourism demand analysis, the decision process and implicit characteristics of the demand can be easily controlled for and studied through the digital tracks left by travellers on the web. The Internet has already assumed a key role in fashions and consumer trends (Horrigan, 2008); holiday preferences revealed by individual searches on an Internet browser provide a new set of empirical evidence, which could furnish real-time information about the main indicators of tourism activities and improve scientific analysis of tourism-demand dynamics.

Since the release of the GT index (in previous years, as well known as Google Insight for Searches), numerous applications have been created to evaluate the information behind the volume of online searches. For instance, Ettredge, Gerdes, and Karuga (2005) focused on the analysis of the unemployment rate. Askitas and Zimmermann (2009) and D'Amuri and Marucci (2009) investigated developing short-term forecasting labour-market models including Google data. Ginsberg et al. (2009) used Google data to detect the flu epidemic. Smith (2012) applied GT indices in modelling the volatility dynamics of financial markets. In the tourism framework, Choi and Varian (2009, 2012) first explored the possibility of adding the Google search index to a simple autoregressive model of tourism arrivals in order to improve the forecasting performance of tourism demand. Recently, Bangwayo-Skeete and Skeete (2015) and Hirashima et al. (2017) implemented Autoregressive-Mixed Data Sampling (AR-MIDAS) regressions to forecast and nowcast tourism demand. Yang, Pan, Evans, and Lv (2015) evaluated the power of GT data and similar search engine volumes in predicting arrivals in China, depicting the importance of reliable web source information. Rivera (2016) defined a dynamic linear model to evaluate the ability to forecast hotel stays in Puerto Rico, considering GT data as the realization of an unobservable process. Dergiades, Mavragani, and Pan (2018) investigated the predictive ability of their corrected Google index (from language bias and platform bias) to analyse causality relationships of the search data and monthly arrivals in Cyprus.

In the present work, we combine the effect of GT data on the accuracy of tourism demand forecasting in neighbourhood small areas with panel data techniques. This class of models has been found relevant to many consumer goods, such as liquor (Baltagi & Griffin, 1995) and cigarettes (Baltagi & Griffin, 2001), but has had very few applications on tourism-demand analysis (Gallego, Rodríguez-Serrano, & Casanueva, 2019; Garín-Muñoz, 2004; Narajan et al., 2010; Sequeira & Maças-Nunes, 2008), especially related to tourism quantities sampled at frequencies higher than quarterly and rarely with a forecasting purpose. A general detailed review of the literature on

panel forecasting models is provided by Baltagi (2007), presenting evidence about the superiority of panel techniques with respect to the time-series approach. Baltagi et al. (2003) found unstable variability of the single regional estimates and worst out-of-sample forecasts. Hoogstrate, Palm, and Pfan (2000) compared forecast accuracy of pooling techniques and single country forecasts, specifically referring to the cases of N fixed and T large. They argued that, when there are similarities among estimated parameters, pooling may improve forecast accuracy. Similar findings have been demonstrated by Gavin and Theodorou (2005) and Fok, Dijk, and Franses (2005) who find that the panel model forecasts of disaggregated series can be more accurate than forecasts of aggregates measures (e.g., total output or unemployment of 48 states).

2.2 | Nowcasting literature

The representativeness and predicting capabilities of GT data can overcome an important problem with the official data: In the real world, most of the quantities and relative measures closely related to tourism and this economic segment are published with a delay that increases when the territorial detail is enhanced. The result is a serious lack of information. In many cases, an evaluation of the forecasting accuracy can only be completed 1 year after the data are published. Fifteen years ago, nowcasting was introduced as a solution to this problem: a new class of econometric methodologies to forecast the timely measure of economic variables. This methodological approach, which takes its name from the combination of the terms now and forecasting (Banbura, Giannone, & Reichlin, 2011), can help mitigate the problems from publication delays in official statistics when measuring tourism activities.

The first contributions to the nowcasting literature were developed in the macroeconomic context (Giannone et al., 2008; Kuzin et al., 2011; Aastveit et al., 2014), with the aim of anticipating variables characterized by low-frequency (e.g., quarterly or yearly data) release. Only recently have the aspects of real-time adjustment and data realization contaminated the tourism framework. This literature provides a set of applications to evaluate nowcast predictions, where time-series models have been augmented by search-volume indices. However, there are a limited number of contributions that address the specific problems related to delay in publication and data vintages. In particular, Camacho and Pacce (2017) considered a dynamic factor model for a real-time database of vintages, finding that the models augmented by Google indices outperform the benchmark models. The authors, in collaboration with Google, built a collection of search-volume index series, tracing the evolution of different tourism queries in real time. They evaluated their nowcasting model referring to short horizons, ranging from 1 to 3 months. This way, the authors completely ignored possible problems caused by publication delays.

The research strategy proposed below overcomes this gap, allowing for different aspects of real-time analysis for local public and private agents: (a) high spatial and frequency detail, (b) a related long delay in publication of official data, and (c) possible substitutive and

complementary effects caused by multiproducts for spatially close destinations.

3 | RESEARCH STRATEGIES

3.1 | Model specification

The dynamic panel data model for the tourism demand for the six neighbourhood territories is defined by

$$Y_{i,t} = \rho_0 + \rho_1 Y_{i,t-1} + \rho_{12} Y_{i,t-12} + D'_{i,t} \delta + \eta_i + \varepsilon_{i,t}, \tag{1}$$

where $i = 1, \dots, N$, $N = 6$, indicates the statistical unit (destinations); and $t = 1, \dots, T$ is the temporal instant at which we evaluate the dependent variables $y_{i,t}$ (logarithm of the arrivals). η_i is the individual specific effect, and $\varepsilon_{i,t}$ is the remainder error; $D_{i,t}$ is a matrix of seasonal dummies (11 dummies, the last variable omitted to avoid multicollinearity problems) introduced to detect deterministic seasonal patterns and furnishing an interpretation of the seasonal fluctuations and their effects on the dependent variable (Shen, Li & Song, 2008). This choice is supported by Canova and Hansen's (1995) test results reported in the last column of Table 1 for seasonality stability of the series. Specifically, the tests do not reject stable seasonal pattern in the series.

In the methodological literature on dynamic panel analysis, the main problem consists in the dependence between the individual effects and the lagged dependent variable $y_{i,t-1}$. In order to overcome the resulting estimation problems, alternative transformations have been proposed in the literature with the aim of evaluating panel dynamic relationships and wiping out the individual effects

(Anderson & Hsiao 1982; Arellano & Bond 1991; Arellano & Bover 1995; Blundell & Bond 1998). In this paper, we consider the difference generalized method of moments (GMM) estimation approach proposed by Arellano (1988) and Arellano and Bond (1991). In addition to having good asymptotic properties,² difference GMM, unlike IV estimators, permits both the use of internal-lagged variables as instruments and increasing efficiency. Furthermore, differently from system GMM, difference GMM is not affected by the likelihood of strong proliferation of instruments and then by overidentification problems (Gallego, Rodríguez-Serrano, & Casanueva, 2019).

To assess the predictive power of GT data, we augment the panel data model introducing the variables $GT'_{i,t-12}$ of Google searches at lag $t - 12$ (after log transformation) to exploit the information from the search engine query for the specified keywords. GT variables are introduced to account for holiday destination preferences as well as for possible structural changes caused by general economic quadrature, word of mouth phenomena, exchange-rate changes, and specific regional situations (e.g., terrorism, wars, and political turbulence). Choosing lag $t - 12$ for both the lagged dependent variable and the explanatory variable, we can consider one of the main characteristics of the tourism demand for Rimini: loyalty. In 2014, the European Union-funded SEE Project for territorial cooperation—InTourAct (Province of Rimini, the University of Bologna, and the Piepoli Institute)—loyal tourists were found to count for 60% of the total, with nearly 90% intending to return (next year).

Starting from the baseline model of Equation (1), the augmented forecasting panel model is given by

$$Y_{i,t} = \rho_0 + \rho_1 Y_{i,t-1} + \rho_{12} Y_{i,t-12} + D'_{i,t} \delta + GT'_{i,t-12} \gamma + \eta_i + \varepsilon_{i,t}. \tag{2}$$

TABLE 1 Descriptive statistics for tourist arrivals in the province, January 2009–December 2016

	Mean	Median	Minimum	Maximum	CV	DH	CH		
Riccione	61,601	42,034	9,582	February 12	173,623	August 15	0.821	57.470***	0.296
Rimini	133,590	96,082	27,248	February 14	361,327	August 15	0.881	51.658***	0.404*
Cattolica	25,111	9,153	1,292	November 16	94,846	August 15	1.033	90.285***	0.119
Misano	10,770	2,363	124	February 13	41,896	August 15	1.035	90.763***	0.139
Bellaria	29,062	15,072	748	February 15	110,686	August 15	1.033	76.546***	0.157
Hinterland	5,056	4,461	1,644	January 09	12,865	August 16	1.152	48.056***	0.213
Province	265,190	167,650	43,487	February 14	794,490	August 16	0.927	63.218***	-
Tests									
Chow (poolability)			3.519***						
Wooldridge ($p = 1$ autocorrelation)			45.911***						
Born–Breitung ($p > 1$ autocorrelation)			12.90***						
LR (heteroskedasticity)			290.05***						
Pedroni (cointegration)	19.29***	-23.99***	-14.31***	-3.24***	-20.07***	-14.37***	1.82**		

Abbreviations: CH, Canova and Hansen test; CV, coefficient of variation; DH, Doornik-Hansen test; LR, likelihood ratio.

***1% significance level.

**5% significance level.

*10% significance level.

FIGURE 1 Province of Rimini

3.2 | The data

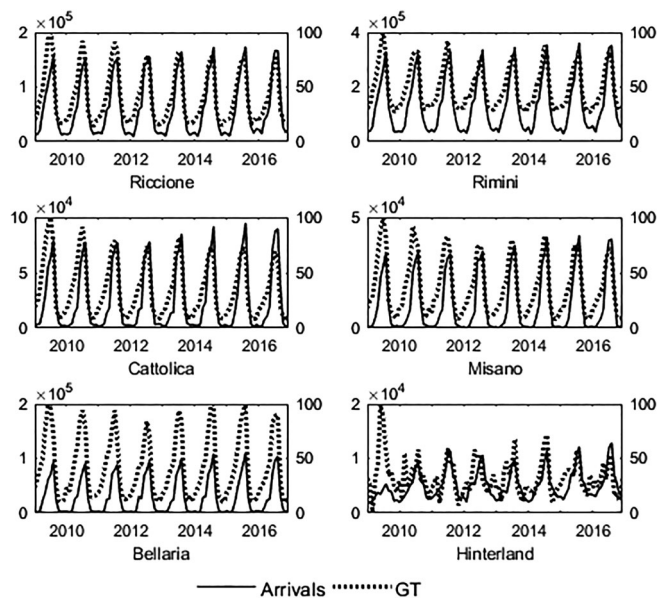
The geographical position and layout (see Figure 1) are defining features of the area around Rimini. The largest municipalities are along the Adriatic Sea coast, and the inland areas are rich in historical attractions, including Roman relics and Mediaeval castles.

Rimini offers a wide set of attractions and tourism products over an area of only 863.6 km². As a result, tourism plays a significant role in the economic balance of the province, accounting for 30% of the total product of the whole area and has the second largest tourism GDP per capita among Italian provinces. In 2016, the international tourism demand in Rimini provided €579 million, with monthly tourism arrivals (total, as the sum of international and Italian) ranging from 55,610 in February to 764,443 in August.

Data on monthly tourism arrivals are obtained from the Emilia Romagna region website (<https://www.regione.emilia-romagna.it>) for the interval January 2009–December 2016. Specifically, we use arrivals as a proxy for the Rimini Province tourism demand. The highest spatial detail for which data are available is defined by six destinations across the territory: Riccione, Rimini, Cattolica, Misano, Bellaria, and the inland areas of the province. In contrast to the first five seaside locations, the inland area has more than one municipality, each offering different magnitudes of attractions but with similar characteristics of tourism supply. As seen in Figure 2, tourism arrivals (solid lines) at each destination show strong summer flows and a weaker flow in the winter.

Table 1 provides some basic statistics of tourism demand, including Chow's (1960) poolability test (significance obtained with bootstrapped critical values provided in Bun, 2004), Wooldridge (2002) and Born and Breitung's (2016) tests for autocorrelation, likelihood ratio test for detecting heteroskedasticity, Canova and Hansen (1995) stability test, and Pedroni's (1999) test for cointegration. All test results support the proposed specification.

Findings on Rimini and Riccione are very similar, especially in the coefficient of variation and the Doornik–Hansen normality test. Moreover, we note that the coastal destinations of Cattolica, Misano,

**FIGURE 2** Arrivals and GT data for the six destinations, from January 2009 to December 2016

and Bellaria are defined by the weakest tourist flows and show closed coefficient of variation and Doornik–Hansen normality values. The inland area presents different results from the other territories, still referring to the instant at which the maximum appears (August 2016). Last column (CH) in Table 1 shows the results from the Canova and Hansen (1995) stability test at Lag 12. For each destination, we do not reject the null of stable seasonality at the 5% significance level.

The search query volumes are obtained from the website <https://trends.google.com/trends/>. Since 2004, Google has provided a search query index that reflects the magnitude of World Wide Web searches for all possible keywords. The Google algorithm provides tracked and categorized searches for a specific query, which includes the searches made around the keywords. This aims to deal with and identify the searches performed using all possible meanings of a specific term.

Through this perspective, GT allows the queries to be refined in two ways: selecting specific categories (all categories, travels, finance, food and drink, etc.) and suggesting the meaning of an inquirer's answers. Hence, for this study, the category of interest is the first important aspect to be defined when downloading the GT.

Even though the literature shows different procedures have been adopted to choose the most informative combination of keywords and categories (for a survey, see Li et al. (2017)), in our study, we refine the searches in the category "Hotel & Accommodation": first, in Rimini, the tourism sector of second homes can be considered marginal with respect to hotel accommodations; and second, the Emilia Romagna region quantifies the arrivals using hotel accommodations.

The search query variables are obtained for the keywords "rimini," "riccione," "cattolica," "misano," "bellaria," and a sixth query built as a composite of the names of several villages in the hinterland inland areas. This last choice basically covers the unavailability of a unique measure of search volume for this territory due to the lack of a specific name that identifies one GT series for the whole destination. We suggest building this search query index with the names of the five main destinations (in terms of arrivals). The final query for the inland area of the province is then given by "san leo" + "santarcangelo di romagna" + "pennabilli" + "verucchio" + "mondaino".

A third aspect that must be considered is the time interval of the downloaded data. Every week, the (absolute) number of searches is obtained referring to a different sample of IP addresses and then is normalized by the total searches of the week. This adjustment provides not only a data update on the maximum of the series (the cases

in which a new maximum of searches is identified or when the set and investigated time interval exclude the observed maximum) but also some small differences among the same series downloaded in two different weeks. In this way, every time we update the training set, we cannot add a new observation to the GT series, but we have to download all of the data again. Both dependent and independent variables are plotted in Figure 2, with solid and dotted lines, respectively. The seasonal pattern of the arrivals is also observed for the GT series, with peaks emerging mostly in July and troughs in winter. Contrary to the other series, the volumes of the searches for inland areas show a not-smoothed seasonal dynamic.

4 | RESULTS

In Table 2, we show baseline and augmented model estimates and tests on linear restrictions and serial correlation of two specific training sets: the first from January 2009 to December 2014 and the last from January 2009 to December 2015 (from left to right). In particular, we refer to the Wald test for $H_{0, b} : \rho_1 = \rho_{12}$ for baseline models cases and $H_{0, a} : \rho_1 = \rho_{12} = \gamma = 0$ for the augmented ones. In the last row of Table 3, we summarize the results of Arellano–Bond test from the second to the 12th-order autocorrelation in first-differenced errors (Arellano & Bond, 1991). Specifically, we refer to the number of times that the tests cannot be rejected at 1% significance level.

Findings suggest that autocorrelation coefficients and the estimates of the GT variables impacts are always positive and significant,

	January 2009–December 2014		January 2009–December 2015	
	Base	Augmented	Base	Augmented
ρ_1	0.235***	0.229***	0.197***	0.171***
ρ_{12}	0.707***	0.694***	0.752***	0.761***
γ		0.125***		0.003**
δ_1	−0.024	−0.026	−0.038	−0.031
δ_2	−0.078	−0.136**	−0.138**	−0.175***
δ_3	0.106	0.129**	0.058	0.074
δ_4	0.369**	0.286***	0.261*	0.179**
δ_5	0.159	0.042	0.151*	0.056
δ_6	0.329**	0.232**	0.236*	0.140
δ_7	0.226**	0.098	0.174*	0.063
δ_8	0.224**	0.116	0.165*	0.067
δ_9	−0.101***	−0.112	−0.105**	−0.113*
δ_{10}	−0.23**	−0.229***	−0.230**	−0.228**
δ_{11}	−0.162***	−0.129**	−0.176***	−0.157***
ρ_0	0.483	0.277***	0.469	0.319***
Wald χ^2_v	3,173.20	7,329.57	2,793.00	9,323.99
Autocorr. Accep.	0	0	0	0

TABLE 2 Estimates of the baseline and the augmented models for the two training sets: January 2009–December 2014 and January 2009–December 2015

***1% significance level.

**5% significance level.

*10% significance level.

TABLE 3 Mean absolute error and the root mean square error indexes calculated on the forecasting periods January–December 2015 and January–December 2016

MAE	January–December 2015		January–December 2016	
	Base	Augmented	Base	Augmented
Riccione	0.1992	0.1948	0.2160	0.1963
Rimini	0.2099	0.1412	0.2649	0.1850
Cattolica	0.1934	0.1898	0.2527	0.2297
Misano	0.1823	0.1569	0.3037	0.2875
Bellaria	0.2426	0.2400	0.2653	0.2671
Hinterland	0.1525	0.1550	0.1315	0.1134
RMSE				
Riccione	0.2707	0.2565	0.2692	0.2398
Rimini	0.2467	0.1694	0.3103	0.2303
Cattolica	0.2444	0.2196	0.3629	0.3436
Misano	0.2675	0.2200	0.4470	0.4265
Bellaria	0.3360	0.3462	0.4509	0.4382
Hinterland	0.1836	0.1858	0.1686	0.1407

Abbreviations: MAE, mean absolute error; RMSE, the root mean square error.

as also confirmed by the Wald tests. The sign of the dummy variables remain constant over the two samples: we observe negative signs for autumn and winter dummies ($\delta_1, \delta_2, \delta_9, \delta_{10}$, and δ_{11}), and positive impacts in spring and summer dummies ($\delta_3, \delta_4, \delta_5, \delta_6, \delta_7$, and δ_8). The 0 values in the last rows guarantee the absence of correlation and so the validity of the moment conditions used by the Arellano–Bond estimator for the four applications.

The forecasting and nowcasting abilities of the two models have been evaluated also referring to further rival specifications. In particular, we consider the univariate time series models defined by

$$Y_{i,t} = \phi_0 + Y_{i,t-12} + \epsilon_{i,t}, \quad (4)$$

for $i = 1, \dots, 6$, and

$$\Delta_{12}Y_{i,t} = \phi_0 + \phi_{12}\Delta_{12}Y_{i,t-12} + \beta\Delta_{12}GT_{i,t-12} + u_{i,t}, \quad (5)$$

where $\Delta_{12}'_{i,t-12} = Y_{i,t} - Y_{i,t-12}$ is the 12 differences of both dependent and independent variables. The third and fourth rival specifications are naïve forecasting models, identified below by N and defined by $\hat{Y}_{i,T+1} = Y_{i,T}$, and Holt–Winters' triple exponential smoothing (following Gunter & Onder, 2015) from the class of exponential smoothing method.

The forecast accuracy measures employed in this work are the mean absolute error (MAE) and the root mean square error (RMSE); as pointed out by Gunter and Onder (2015), the characteristic of these quantities, calculated on the logarithmic series, is to approximate the related percentage indices (mean absolute percent error and root mean square percent error) of the untransformed variables. We also consider a third measure of accuracy, built as a ratio of the absolute

errors. Specifically, the relative quantities are defined by the ratios of the absolute errors done with the benchmark models relative to the absolute error obtained with the augmented model, specifically defined by

$$R_{T+h} = \frac{|Y_{T+h} - \hat{Y}_{T+h,b}|}{|Y_{T+h} - \hat{Y}_{T+h,a}|}, \quad h = 1, \dots, 12, \quad (6)$$

where Y_{T+h} is the variable observed h steps forward the end of the estimation period T ; and $\hat{Y}_{T+h,b}$ and $\hat{Y}_{T+h,a}$ are the forecasts for the benchmark and the augmented models, respectively. Through these quantities, we obtain a direct measure of comparison because values equal to 1 represent equal errors for the two models, $R_{T+h} > 1$ assesses a better performance of the benchmark model, and $R_{T+h} < 1$ confirms the goodness of the augmented forecasting model.

Moreover, the lack of information caused by the publication delay for official data aggravates the difficulty of obtaining reliable out-of-sample results. Thus we further implement a nowcasting application. We collect the GT variables raising alternative samples of search-volume data: specifically, we start from the period January 2009–May 2016, and we add 1 month to this sample, until the interval January 2009–September 2016, obtaining a different series for the search query indices for each interval. To enhance the characteristic of reliability of the nowcasting application, we evaluate horizons $h \geq 8$ in line with availability of high (temporal and spatial) detailed official data, where classical delays range from 8 to 15 months.

To summarize, we consider two specific forecast evaluation schemes: the first one general to evaluate the accuracy of the whole yearly performances of each destination; and second, the prediction of summer months when they appear as 1, 2, ..., 12 steps ahead. For the nowcasting scheme, first we focus on the predicting capability of the model measuring MAEs and RMSEs for the period May–September 2016; hence, we investigate the nowcasting performances of the rivals in summer months of June, July, and August when they appear as horizons $h = 8, 10, 12$, and 15.

4.1 | Forecasting.1

The first training set is defined by January 2009–December 2014 ($T = 60$), and the second training sample is obtained adding 12 months to the previous interval (until December 2015, $T = 72$). In Table 3, we evaluate the forecast performances of the benchmark and augmented models through MAEs and RMSEs for the following 12 months of 2015 and 2016, respectively.

Even if the magnitude of the errors is quite strong in both training sets (MAEs ranging from 0.1315 to 0.3037 for the benchmark and from 0.1134 to 0.2875 for the augmented model; benchmark RMSEs vary between 0.1686 and 0.4509 and 0.1407 and 0.4382 for the augmented specification), the augmented model generally produces smaller errors than the baselines (in bold). For the forecasting period January–December 2016 in particular (fourth and fifth columns), we observe five augmented model MAEs and all the RMSEs, lower than

TABLE 4 Comparison of the summer forecasting performances of the benchmark and the augmented model through R_{T+h} , and mean absolute error values for the six rival models

	h	Riccione	Rimini	Cattolica	Misano	Bellaria	Hinterland
June	1	1.19	1.22	1.01	0.85	0.99	0.60
	2	0.86	1.33	0.91	0.45	1.29	0.72
	3	0.99	1.46	0.92	0.40	1.22	0.73
	4	1.05	1.45	1.23	1.23	2.02	0.42
	5	1.02	1.43	1.23	1.31	2.01	0.44
	6	1.01	1.07	1.18	1.70	2.15	0.40
	7	0.58	0.74	0.93	0.79	0.18	0.47
	8	0.50	0.72	0.61	0.23	0.06	1.18
	9	0.30	0.52	0.99	1.13	0.18	0.46
	10	0.32	0.53	0.99	1.12	0.18	0.46
	11	0.32	0.53	0.99	1.13	0.18	0.46
	12	0.33	0.54	0.99	1.13	0.16	0.46
MAE _g		0.05	0.11	0.11	0.06	0.03	0.12
MAE _b		0.07	0.13	0.12	0.07	0.04	0.22
MAE _{RW}		0.06	0.12	0.13	0.09	0.07	0.02
MAE _{AR+GT}		0.07	0.10	0.12	0.19	0.05	0.07
MAE _N		0.49	0.42	0.99	0.92	0.71	0.15
MAE _{ETS}		0.25	0.14	0.19	0.07	0.22	0.18
July	1	1.02	0.92	1.31	2.14	0.67	2.59
	2	0.92	0.86	1.17	1.46	0.84	0.38
	3	0.94	0.89	1.10	0.57	0.58	0.80
	4	0.93	0.89	1.11	0.54	0.58	0.78
	5	0.69	0.68	1.00	0.85	0.38	0.73
	6	0.73	0.71	1.03	0.92	0.39	0.57
	7	0.96	0.86	1.18	1.02	0.50	0.51
	8	0.80	0.85	0.99	1.19	0.49	0.75
	9	2.72	1.57	2.60	2.13	2.37	0.35
	10	0.65	0.74	1.01	1.60	0.48	0.94
	11	0.66	0.75	1.01	1.63	0.48	0.85
	12	0.65	0.74	1.01	1.62	0.47	0.80
MAE _g		0.09	0.16	0.09	0.04	0.04	0.04
MAE _b		0.10	0.19	0.08	0.03	0.07	0.06
MAE _{RW}		0.06	0.09	0.14	0.09	0.10	0.06
MAE _{AR+GT}		0.08	0.10	0.15	0.20	0.12	0.12
MAE _N		0.17	0.31	0.19	0.26	0.28	0.45
MAE _{ETS}		0.27	0.15	0.21	0.09	0.20	0.13
August	1	0.42	0.65	1.00	1.05	1.36	0.80
	2	0.85	0.84	0.55	0.72	0.87	0.41
	3	0.95	0.80	0.51	0.73	0.99	0.41
	4	0.59	0.84	1.24	0.95	0.90	0.71
	5	0.59	0.83	1.23	0.94	0.90	0.70
	6	1.27	0.45	1.43	0.90	1.27	0.44
	7	1.35	0.48	1.45	0.90	1.25	0.46
	8	2.04	0.77	1.77	0.84	1.42	0.50
	9	1.70	0.37	1.09	0.95	2.04	0.42

(Continues)

TABLE 4 (Continued)

h	Riccione	Rimini	Cattolica	Misano	Bellaria	Hinterland
10	0.22	1.84	0.71	0.89	1.11	0.78
11	3.06	0.13	1.13	0.86	2.26	0.24
12	3.24	0.16	1.12	0.86	2.28	0.25
MAE_G	0.03	0.06	0.06	0.09	0.07	0.07
MAE_B	0.03	0.09	0.06	0.10	0.05	0.13
MAE_{RW}	0.07	0.03	0.04	0.07	0.06	0.01
MAE_{AR+GT}	0.07	0.03	0.01	0.03	0.06	0.03
MAE_N	0.06	0.09	0.11	0.09	0.15	0.10
MAE_{ETS}	0.35	0.18	0.17	0.09	0.35	0.22

Abbreviation: MAE, mean absolute error.

TABLE 5 Mean absolute error and the root mean square error indexes calculated on the nowcasting period May 2016–September 2016

MAE	$h = 8$		$h = 10$		$h = 12$		$h = 15$	
	Base	Augm.	Base	Augm.	Base	Augm.	Base	Augm.
Riccione	0.1380	0.1323	0.1325	0.1151	0.1271	0.1077	0.1201	0.1849
Rimini	0.1726	0.1365	0.1562	0.1228	0.1667	0.1273	0.1636	0.2049
Cattolica	0.1570	0.1519	0.1513	0.1436	0.1448	0.1382	0.1368	0.1449
Misano	0.1469	0.1307	0.1365	0.1412	0.1315	0.1409	0.1204	0.0979
Bellaria	0.1452	0.1400	0.1374	0.1288	0.1292	0.1164	0.1207	0.0463
Hinterland	0.1418	0.1159	0.1676	0.1407	0.1648	0.1578	0.1605	0.2240
RMSE								
Riccione	0.1604	0.1502	0.1532	0.1407	0.1526	0.1314	0.1408	0.2120
Rimini	0.1811	0.1516	0.1684	0.1319	0.1748	0.1308	0.1748	0.2074
Cattolica	0.1986	0.1840	0.1905	0.1802	0.1872	0.1705	0.1682	0.1755
Misano	0.2243	0.1994	0.1902	0.2030	0.1916	0.1966	0.1691	0.1165
Bellaria	0.2935	0.2208	0.2612	0.2173	0.2640	0.2025	0.1889	0.0537
Hinterland	0.2515	0.1721	0.2388	0.1822	0.2398	0.1875	0.2030	0.2500

Abbreviations: MAE, mean absolute error; RMSE, the root mean square error.

their benchmark counterparts, highlighting a general performance enhancement after inclusion of the GT variables.

4.2 | Forecasting.2

Given the seasonal nature of arrivals in the province, we propose further considerations. In particular, we implement a recursive procedure to forecast $h = 1, \dots, 12$ horizons, starting from the estimation sample January 2009–December 2014, and we focus on the predicted values for summer (June, July, and August), differentiating by the horizon at which these months appear as forecasts. Even if the number of training sets, and hence forecasts, is small, several results and implications can be derived.

For each month we report in Table 4, the ratio R_{T+h} of Equation (5) for the baseline and the augmented models and the mean of the 12 horizons' absolute errors for all rival specifications, that is, MAE_B is the mean of the absolute errors with the benchmark panel

specification; MAE_G refers to the absolute errors of the augmented models; MAE_{RW} is calculated for the forecasts obtained through the specification in Equation (3); MAE_{AR+GT} for Equation (4); and MAE_N and MAE_{ETS} for naïve and exponential smoothing model, respectively. If we compare MAEs in Table 3 and MAEs in Table 4, we see that both the baseline and the augmented model provide forecasts that are more accurate in summer than in the other periods of the year. Referring to the R_{T+h} quantities, most of the ratios in June and July are lower than 1, assessing the forecasting ability of the augmented model.

Focusing on each destination, the main advantages including GT variables can be noted for the inland area and Rimini, especially in July and August. A different result is obtained for Cattolica (for which only 12 R_{T+h} ratios over 36 are lower than 1) and for Misano, the only bathing destination close to Cattolica in the province. The worst ratios are shown for Riccione when August represents horizons $h = 11$ and $h = 12$. However, if we observe the terms of these relative measures, we show that the R_{T+h} are given by

$$R_{T+11} = \frac{|y_{T+11} - \hat{y}_{T+11,b}|}{|y_{T+11} - \hat{y}_{T+11,d}|} = \frac{0.015}{0.047}, R_{T+12} = \frac{0.014}{0.044},$$

highlighting the goodness of the augmented model performances with MAE indices less than 0.05 (representing approximately a mean absolute percent error of 5% of the arrivals).

4.3 | Nowcasting.1

In the nowcasting application, we focus on the predictions of arrivals in May, June, July, August, and September 2016, only when these months appear as $h = 8, 10, 12,$ and 15 for the best four rival models identified in Table 4—the benchmark panel specification, the augmented model, the random walk specification, and the Autoregressive (AR) model augmented by GT. In Table 5, we show the MAE and the RMSE indices of the 5-month forecasts evaluated at each horizon.

For $h = 8$, we confirm an enhancement of the results for all the destinations after introducing the GT variables. In the cases of $h = 10$ and $h = 12$, only Misano exhibits better results with the baseline model. For the largest horizon, $h = 15$, we find the performance overturned, all MAEs of the benchmark are lower than MAEs of augmented specifications despite the cases of Misano and Bellaria.

4.4 | Nowcasting.2

We show results for summer in Table 6, through R_{T+h} ratios for the best four rival models identified in Table 4—the benchmark panel specification, the augmented model, the random walk specification, and the AR model augmented by GT. In this case, no one model clearly outperforms the others.

Considering the first rival, that is, the baseline panel model, the main advantages in forecast accuracy can be noted for horizons $h = 10,$ and 12 , where most of ratios are less than 1. The worst

h = 8	Base			RW			AR + GT		
	June	July	August	June	July	August	June	July	August
Riccione	0.51	1.18	2.88	1.02	1.71	0.76	0.90	1.24	0.74
Rimini	0.71	1.08	0.15	1.18	2.13	0.40	1.42	2.21	0.34
Cattolica	0.54	1.03	1.26	0.81	0.71	2.30	1.17	0.70	2.13
Misano	0.14	0.83	1.06	0.20	0.39	1.99	0.25	0.17	2.77
Bellaria	0.04	0.70	2.17	0.05	0.54	2.07	0.25	0.44	1.43
Hinterland	1.27	0.12	0.29	9.19	0.08	6.16	2.88	0.04	0.92
h = 10									
Riccione	0.97	0.73	0.24	1.03	1.05	0.20	0.90	0.74	0.20
Rimini	1.03	0.70	0.67	1.27	1.45	1.25	1.44	1.49	1.23
Cattolica	0.82	0.83	0.64	0.83	0.61	1.58	1.17	0.61	1.39
Misano	0.29	1.10	0.79	0.20	0.50	1.78	0.22	0.22	2.30
Bellaria	0.17	0.30	1.31	0.08	0.23	2.24	0.39	0.19	1.72
Hinterland	0.90	1.38	0.17	9.71	1.25	4.18	3.09	0.68	0.68
h = 12									
Riccione	0.98	0.73	0.04	1.05	1.06	0.01	0.91	0.75	0.01
Rimini	1.04	0.71	0.95	1.22	1.61	2.70	1.43	1.53	3.57
Cattolica	0.82	0.82	0.95	0.82	0.59	1.64	1.12	0.59	1.45
Misano	0.26	1.08	1.12	0.18	0.47	2.10	0.19	0.20	2.58
Bellaria	0.18	0.28	1.63	0.08	0.22	1.77	0.37	0.18	1.28
Hinterland	0.90	1.24	0.80	9.50	1.22	20.54	2.96	0.65	2.95
h = 15									
Riccione	1.91	1.52	5.77	2.22	1.08	0.90	1.40	1.05	1.37
Rimini	1.34	1.08	2.26	2.40	1.32	31.72	2.82	1.51	16.68
Cattolica	1.25	1.58	0.50	1.16	0.68	1.23	1.67	0.81	1.47
Misano	0.25	2.10	1.72	0.56	0.74	2.86	0.17	0.28	1.52
Bellaria	0.88	0.84	1.55	0.53	0.39	18.74	1.23	0.34	1.34
Hinterland	1.54	1.68	1.63	5.77	0.71	22.07	2.72	0.38	2.14

TABLE 6 Comparison of the summer nowcasting performances of the best four rival models through R_{T+h}

Abbreviations: AR, autoregressive; GT, Google Trends; RW, random walk.

situation is represented by $h = 15$. If we consider the univariate time-series models, the predictive power of the augmented panel specification is not so clear. However, we can find that the enhancement of accuracy is obtained for destinations representative of similar products and with the weakest tourist flows among the six territories: Cattolica, Misano, and Bellaria. As for forecasting results, the summer evaluation in Table 6 shows errors that are quite different from the mean obtained in Table 5, finding the goodness of the model, and therefore of the GT data, to capture the summer dynamic.

From a practitioner point of view, it is extremely important for public and private agents operating at different spatial detail to understand, model, and accurately predict the number of tourists coming to the destination. The aims are to prevent and sustain low-level demand conditions (e.g., in winter months and local inefficiency of smaller destinations, this one often characterized by a low level of attractiveness) while planning and programming both an adequate number of available accommodations and employees. Accurately predicting the specific-municipality demand provides the opportunity to track the performance of competing destinations (as in the case of $h = 15$ for Cattolica, Misano, and Bellaria), the opportunity for minor destinations (inland areas and Misano) to monitor the role of events in neighbourhood municipalities in the off-peak as well as peak season, and the opportunity to detect the impact of shocks at different micro-economic levels on tourist arrivals in real time through contractions in search volumes. These issues also suggest that greater access to tourism destinations could be improved by means of greater visibility and a more accurate and proper marketing strategy on the Internet, especially for small isolated municipalities with small magnitude attractions (e.g., inland areas of the province). All these practical aspects require opportune and prompt predictions of tourist flows dealing with data availability and publication delays of official data. Despite the importance, both availability and delay are usually disregarded in tourism literature often leading researchers to highlight accuracy of inappropriate and almost unrealistic forecasting schemes.

5 | CONCLUSIONS AND IMPLICATIONS

Short- and long-term predictions provide answers to different questions of economic agents. Frechtling (2001) summarized the relationship between tourism managerial requirements and the temporal intervals to adopt efficient and effective economic, managerial, and marketing actions. For local destinations, the gap in official data, which affects either short- and long-term predictions, becomes dramatic, strongly undermining the efficiency of the decision-making processes. However, small towns where the tourism industry is at the core of the local economy may benefit from an accurate analysis of the dynamics of the tourism flows to a larger extent than the metropolitan areas. The aim of this paper is to respond to this shortcoming by combining relevant concepts introduced (separately) in the tourism framework over the past decade.

First, the use of a panel specification to model and forecast tourism demand of both complementary and substitutive neighbourhoods.

Second, the use of GT indices to enhance the predictive capability of forecasting and nowcasting models. This strand of statistical and econometric literature has been recently introduced to define models able to predict the present, the very near future, and the very recent past. Researchers either become interested in building specific instruments (for a review, see Banbura, Giannone, & Reichlin (2011)) able to deal with mixed frequency data or augmenting forecasting models with ex ante measures of the interesting variables. Following the latter direction, tourism literature focuses on using advanced techniques to deal with models augmented by online search data, usually avoiding important aspects of forecasting and nowcasting approaches. The third and fourth relevant concepts are found in two important and ignored problems of local economic agents: the availability of official data and the spatial detail of the quantities involved in the analysis. Even if results differ across municipalities and forecast horizons (Witt & Witt, 1995), focusing on weak summer tourism flows and adopting a more realistic forecasting scheme, the augmented panel data model proposed in this paper provides more accurate forecasts than the baseline counterpart and four rival univariate specifications, which underscore the importance of jointly modelling weak and strong tourism demand for neighbouring destinations at monthly frequency.

Despite the advantages identified with this approach, limitations and further developments can be illustrated. First, the combination of temporal availability and the model specification: in fact, due to the size of the data set joint with the number of parameters in the panel specification, we avoid increasing the number of training sets, taking care of the general advantages of panel-data techniques. Further developments should consider increasing the temporal dimension and/or the number of destinations. Second, given the geographic proximity of the analysed municipalities, the analysis would be improved by increasing the data set from a spatial point of view,³ specifically increasing the number of neighbouring territories and providing a spatiotemporal panel data investigation.

ORCID

Silvia Emili  <https://orcid.org/0000-0002-7086-0650>

ENDNOTES

¹ From the Italian National Institute of Statistics available at <http://dati.istat.it/>.

² In our case, $N = 6$ and $T > 60$. As pointed out by Alvarez and Arellano (2003), these generalized method of moment panel data estimators are asymptotically normal distributed and unbiased if the number of individuals N is fixed and T is infinity.

³ We thank an anonymous referee for this suggestion.

REFERENCES

- Aastveit, K., Gerdrup, K., Jore, A., & Thorsrud, L. (2014). Nowcasting GDP in real time: A density combination approach. *Journal of Business and Economic Statistics*, 32, 48–68.
- Alvarez, J., & Arellano, M. (2003). The time series and cross-section asymptotics of dynamic panel data estimators. *Econometrica*, 71(4), 1121–1159.

- Anderson, T., & Hsiao, C. (1982). Formulation and estimation of dynamic models using panel data. *Journal of Econometrics*, 18, 47–82.
- Arellano, M. (1988). *An alternative transformation for fixed effects models with predetermined variables*. Applied economics discussion paper no. 57. Institute of Economics and Statistics, University of Oxford.
- Arellano, M., & Bond, S. (1991). Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *Review of Economic Studies*, 58, 277–297.
- Arellano, M., & Bover, O. (1995). Another look at the instrumental variables estimation of error component models. *Journal of Econometrics*, 68, 29–51.
- Askatas, N., & Zimmermann, K. F. (2009). Google econometrics and unemployment forecasting. *Applied Economics Quarterly*, 55(2), 107–120.
- Athanasopoulos, G., Hyndman, R. J., Song, H., & Wu, D. C. (2011). The tourism forecasting competition. *International Journal of Forecasting*, 27, 822–844.
- Baltagi, B. H. (2007). Forecasting with panel data. *Center for Policy Research Working Paper No. 91*. Available at <http://dx.doi.org/10.2139/ssrn.1815319>
- Baltagi, B. H., Bresson, G., Griffin, J. M., & Piroette, A. (2003). Homogeneous, heterogeneous or shrink age estimators? Some empirical evidence from French regional gasoline consumption. *Empirical Economics*, 28, 795–811.
- Baltagi, B. H., & Griffin, J. M. (1995). A dynamic demand model for liquor: The case for pooling. *Review of Economics and Statistics*, 77, 545–553.
- Baltagi, B. H., & Griffin, J. M. (2001). The econometrics of rational addiction: The case of cigarettes. *Journal of Business and Economic Statistics*, 19, 449–454.
- Banbura, M., Giannone, D., & Reichlin, L. (2011). Nowcasting. In M. P. Clements & D. F. Hendry (Eds.), *The Oxford handbook of economic forecasting* (pp 63–90). Oxford University Press.
- Bangwayo-Skeete, P. F., & Skeete, R. W. (2015). Can Google data improve the forecasting performance of tourist arrivals? Mixed-data sampling approach. *Tourism Management*, 46, 454–464.
- Blundell, R., & Bond, S. (1998). Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics*, 87, 115–143.
- Born, B., & Breitung, J. (2016). Testing for serial correlation in fixed-effects panel data models. *Econometric Reviews*, 35(7), 1290–1316.
- Bun, M. J. G. (2004). Testing poolability in a system of dynamic regressions with nonspherical disturbances. *Empirical Economics*, 29(1), 89–106. <https://doi.org/10.1007/s00181-003-0191-3>
- Camacho, M., & Pácces, M. J. (2017). Forecasting travellers in Spain with Google's search volume indices. *Tourism Economics*, 24, 434–448. <https://doi.org/10.1177/1354816617737227>
- Canova, F., & Hansen, B. E. (1995). Are seasonal patterns constant over time? A test for seasonal stability. *Journal of Business and Economic Statistics*, 13(3), 237–252.
- Chen, C. M., Lai, F. T., & Hua, K. T. (2019). A linkage between internet use and tourism activities. *Current Issues in Tourism*, 22(3), 291–300. <https://doi.org/10.1080/13683500.2017.1403417>
- Choi, H., & Varian, H. (2009). *Predicting the present with Google Trends*. Economics Research Group, Google Inc., California, USA.
- Choi, H., & Varian, H. (2012). Predicting the present with google trends. *Economic Record*, 88(1), 2–9.
- Chow, G. C. (1960). Tests of equality between sets of coefficients in two linear regressions. *Econometrica*, 28(3), 591–605. <https://doi.org/10.2307/1910133>
- D'Amuri, F., & Marcucci, J. (2009). Google it! Forecasting the US unemployment rate with a Google job search index. ISER Working paper series pp. 2009–32.
- De Mauro, A., Greco, M., & Grimaldi, M. (2015). What is big data? A consensual definition and a review of key research topics. *AIP Conference Proceedings*, 1644(1), 97–104.
- Dergiades, T., Mavragani, E., & Pan, B. (2018). Google trends and tourists' arrivals: Emerging biases and proposed corrections. *Tourism Management*, 66, 108–120.
- Emili, S., Figini, P., & Guizzardi, A. (2019). Modelling international monthly tourism demand at the micro destination level with climate indicators and web-traffic data. *Tourism Economics*. <https://doi.org/10.1177/1354816619867804>
- Ettredge, M., Gerdes, J., & Karuga, G. (2005). Using web-based search data to predict macroeconomic statistics. *Communications of the ACM*, 48 (11), 87–92.
- Fok, D., van Dijk, D., & Franses, P. H. (2005). Forecasting aggregates using panels of nonlinear time series. *International Journal of Forecasting*, 21, 785–794.
- Frechtling, D. C. (2001). *Forecasting tourism demand: Methods and strategies*. Burlington, MA: Butterworth Heinemann.
- Gallego, A., Rodriguez-Serrano, M. A., & Casanueva, C. (2019). Dynamic panel data models in tourism. *Current Issues in Tourism*, 22 (4), 379–399. <https://doi.org/10.1080/13683500.2018.1467386>
- Garin-Muñoz, T. (2004). Madrid as a tourist destination: Analysis and modelization of inbound tourism. *International Journal of Tourism Research*, 6(4), 289–302. <https://doi.org/10.1002/jtr.489>
- Gavin, W., & Theodorou, A. T. (2005). A common model approach to macroeconomics: Using panel data to reduce sampling error. *Journal of Forecasting*, 24, 203–219.
- Giannone, D., Reichlin, L., & Small, D. (2008). Nowcasting GDP and inflation: The real time informational content of macroeconomic data releases. *Journal of Monetary Economics*, 55(4), 665–676.
- Gil-Alana, L. A. (2005). Modelling international monthly arrivals using seasonal univariate long memory processes. *Tourism Management*, 26, 867–878.
- Ginsberg, J., Mohebbi, M. H., Patel, R. S., Brammer, L., Smolinski, M. S., & Brilliant, L. (2009). Detecting influenza epidemics using search engine query data. *Nature*, 457(7), 1012–1014.
- Goffi, G., & Cucculelli, M. (2019). Explaining tourism competitiveness in small and medium destinations: The Italian case. *Current Issues in Tourism*, 22(17), 2109–2139. <https://doi.org/10.1080/13683500.2017.1421620>
- Gunter, U., & Onder, I. (2015). Forecasting international city tourism demand for Paris: Accuracy of uni- and multivariate models employing monthly data. *Tourism Management*, 46, 123–135.
- Gunter, U., & Onder, I. (2016). Forecasting city arrivals with Google analytics. *Annals of Tourism Research*, 61, 199–212.
- Hassani, H., & Silva, E.S. (2015). Forecasting with Big Data: A Review. *Annals of Data Science*, 2, 5–19.
- Hirashima, A., Carl, J. J., Bonham, S., & Fuleky, P. (2017). Forecasting in a mixed up world: Nowcasting Hawaii tourism. *Annals of Tourism Research*, 63, 191–202.
- Hoogstrate, A. J., Palm, F. C., & Pfann, G. A. (2000). Pooling in dynamic panel-data models: An application to forecasting GDP growth rates. *Journal of Business Economic and Statistics*, 18, 274–283.
- Horrigan, J. B. (2008). *Online shopping*. Washington, DC: Pew Internet Life and American Project Report.
- Hyndman, R. J., Koehler, A. B., Snyder, R. D., & Grose, S. (2002). A state space framework for automatic forecasting using exponential smoothing methods. *International Journal of Forecasting*, 18, 439–454.
- Jackman, M., & Naitram, S. (2015). Nowcasting tourist arrivals to Barbados. Just Google it! *Tourism Economics*, 21, 1309–1313.
- Kuzin, V., Marcellino, M., & Schumacher, C. (2011). MIDAS vs. mixed-frequency VAR: Nowcasting GDP in the euro area. *International Journal of Forecasting*, 27(2), 529–542.
- Li, X., Pan, B., Law, R., & Huang, X. (2017). Forecasting tourism demand with composite search index. *Tourism Management*, 59, 57–66.
- Lim, C., & McAleer, M. (2001). Cointegration analysis of quarterly tourism demand by Hong Kong and Singapore for Australia. *Applied Economics*, 33(12), 1599–1619.

- Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Byers, A. H. (2011). *Big data: The next frontier for innovation, competition, and productivity*. San Francisco, CA: McKinsey Global Institute. http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation
- Narajan, P. K., Narayan, S., Prasad, A., & Prasad, B. C. (2010). Tourism and economic growth: A panel data analysis for Pacific Island countries. *Tourism Economics*, 16(1), 169–183.
- Pedroni, P. (1999). Critical values for cointegration tests in heterogeneous panels with multiple regressors. *Oxford Bulletin of Economics and Statistics*, 61, 653–670.
- Rivera, R. (2016). A dynamic linear model to forecast hotel registrations in Puerto Rico using Google Trends data. *Tourism Management*, 57, 12–20.
- Rossellò-Nadal, J. (2001). Forecasting turning points in international visitor arrivals in the Balearic Islands. *Tourism Economics*, 7, 365–380.
- Sequeira, T. N., & Maças-Nunes, P. (2008). Does tourism influence economic growth? A Dynamic Panel Data Approach. *Applied Economics*, 40(18), 2431–2441.
- Smith, G. P. (2012). Google Internet search activity and volatility prediction in the market for foreign currency. *Finance Research Letters*, 9(2), 103–110.
- Shen, S., Li, G., & Song, H. (2008). Effect of seasonality treatment on the forecasting performance of tourism demand models. *Tourism Economics*, 15(4), 693–708.
- Song, H., & Li, G. (2008). Tourism demand modelling and forecasting-A review of recent research. *Tourism Management*, 29, 203–220.
- Song, H., & Liu, H. (2017). Predicting tourist demand using big data, in X. Z. & F. D., Eds. *Analytics in smart tourism design*. Tourism on the Verge. Cham: Springer International Publishing Switzerland.
- Song, H., Witt, S. F., & Li, G. (2009). *The advanced econometrics of tourism demand*. New York, London: Routledge.
- Uysal, M., & O'Leary, J. (1986). A canonical analysis of international tourism demand. *Annals of Tourism Research*, 13(4), 651–655.
- Varian, H. (2014). Big data: New tricks for econometrics. *Journal of Economic Perspectives*, 28(2), 3–28.
- Witt, S., & Witt, C. (1991). Tourism forecasting: Error magnitude, direction of change error and trend change error. *Journal of Travel Research*, 30, 26–33.
- Witt, S. F., & Witt, C. A. (1995). Forecasting tourism demand: A review of empirical research. *International Journal of Forecasting*, 11, 447–475.
- Wooldridge, J. M. (2002). *Econometric analysis of cross section and panel data*. Cambridge, MA: MIT Press.
- Yang, X., Pan, B., Evans, J. A., & Lv, B. (2015). Forecasting Chinese tourist volume with search engine data. *Tourism Management*, 46, 386–397.

How to cite this article: Emili S, Gardini A, Foscolo E. High spatial and temporal detail in timely prediction of tourism demand. *Int J Tourism Res*. 2020;22:451–463. <https://doi.org/10.1002/jtr.2348>