Google Trends data and transfer function models to predict tourism demand in Italy

Giovanni De Luca and Monica Rosciano

Abstract

Purpose – The tourist industry has to adopt a big data-driven foresight approach to enhance decision-making in a post-COVID international landscape still marked by significant uncertainty and in which some megatrends have the potential to reshape society in the next decades. This paper, considering the opportunity offered by the application of the quantitative analysis on internet new data sources, proposes a prediction method using Google Trends data based on an estimated transfer function model.

Design/methodology/approach – The paper uses the time-series methods to model and predict Google Trends data. A transfer function model is used to transform the prediction of Google Trends data into predictions of tourist arrivals. It predicts the United States tourism demand in Italy.

Findings – The results highlight the potential expressed by the use of big data-driven foresight approach. Applying a transfer function model on internet search data, timely forecasts of tourism flows are obtained. The two scenarios emerged can be used in tourism stakeholders’ decision-making process. In a future perspective, the methodological path could be applied to other tourism origin markets, to other internet search engine or other socioeconomic and environmental contexts.

Originality/value – The study raises awareness of foresight literacy in the tourism sector. Secondly, it complements the research on tourism demand forecasting by evaluating the performance of quantitative forecasting techniques on new data sources. Thirdly, it is the first paper that makes the United States arrival predictions in Italy. Finally, the findings provide immediate valuable information to tourism stakeholders that could be used to make decisions.

Keywords Time-series analysis, Forecast tourism demand, Big data, Google Trends data, Transfer function model, Foresight approach

1. Introduction

As a result of the COVID-19 epidemic, which has caused a worldwide health, social and economic emergency, travel and tourism are among the most adversely affected industries, with planes grounded, hotels shuttered and travel restrictions implemented in almost every nation on earth. The number of foreign visitors in 2021 compared to 2019 is drastically down: 73% from Africa, 60% from the Americas, 94% from Asia and the Pacific, 62% from Europe and 75% from the Middle East (Unwto, 2022). The epidemic has highlighted the weaknesses of the current capabilities in the tourism industry when it comes to dealing with the fallout from a crisis of this magnitude and reviving the industry.

Tourist destinations need to seek a new normal and find new balances. In having to adapt to periods of uncertainty, in looking for alternative sources of financing, in repositioning and diversifying markets, products and services, it becomes not only essential to restart but also do it so in a more inclusive, innovative and sustainable way (Unwto, 2021). In this context, a notable contribution comes from the analysis of the megatrends such as the evolving visitor demand, the sustainable tourism growth, the enabling technologies and travel mobility. It is needed to adapt...
tourism in an increasingly fast-moving, inter-connected and dynamic environment following a strategic foresight approach (OECD, 2018). It is highly desirable to establish a new governance model between the Institutions, strengthening the organizational, financial and human resources and promoting a data-driven decision process. Good policy and planning decisions are made with authoritative, reliable and accessible data (WTTC, 2022). Recent tourism data indicate that these transformations in the tourism sector must take place in a timely manner not only to adequately address negative shocks but also unsustainable over-tourism phenomena already detected in some locations such as Croatia, Montenegro and Italy in last years (Knezevic et al., 2021). The foresight approach helps tourism to take proactive measures to reduce uncertainty, turbulence and ambiguity. Managerial interest in accurate forecasts of future events and changes in tourism demand is related to the extent to which actions can be implemented to influence the demand, monitor fluctuations for planning or resource capacities as well as manage the (for example, socioeconomic) effects of demand. Hence, the success of tourism businesses largely depends on the ability to predict tourism demand most accurately (Hopken et al., 2019) and timely.

Additionally, given that search queries are increasingly seen as a contemporaneous snapshot of collective consciousness due to the increased usage of the internet for news and information in recent years, what people search for is a good indicator of what they will do in the future (Goel et al., 2010). As a result, information from internet searches is recognized as a significant source of linked contemporaneous data (Jackman and Naitram, 2015).

Besides, emerging tech-savvy generations will become increasingly dominant consumers in the tourism market and recent studies also indicate that post-COVID travelers would have an even more digital profile and explore vacation alternatives on internet platforms. Digital solutions that deliver customized and real-time information will be more appealing to tourists (Knezevic et al., 2021).

The above discussion drivers our motivation to answer the following research question: “How to obtain future scenarios using real-time information provision.” The pandemic strengthens the need for more integrated, timely, comparable and granular data. Improving the evidence base, looking beyond traditional survey approaches using alternate and complementary data sources, to inform policy and business decisions is a key policy priority (OECD, 2022).

Therefore, this paper, following a big data-driven foresight approach (Saritas et al., 2022), predicts international tourism demand using a transfer function model on Google Trends data. Google Trends has become a very popular data source among researchers of a wide variety of fields over the last decade, despite some critical issues in quality (Cebrián and Domenech, 2023).

The work focused on international tourism considering that although domestic travel has been a key factor in the recovery to date and remains essential (Aldao et al., 2022), the full recovery of the global economy will require the full return of international travel (WTTC, 2022).

In detail, it focuses on the US tourism demand in Italy, a country in which the total contribution of tourism and travel to GDP is 13%, to employment is 15% and foreign visitor spend is 8% of total exports (WTTC, 2022). International tourism accounts for 49% in terms of arrivals and the United States of America are the main extra-European market. Americans represent 9% of international arrivals and 30% of extra-European arrivals (Istat, 2019). The US tourist expenditure represents 13% (Bank of Italy, 2020).

Additionally, a rise in interest in national tourism attractions was seen in Italy between January and April 2021, with a proportion of searches for these locations increasing from 55% to 79% (Knezevic et al., 2021). For the public and private operators of the Italy tourist destination, the availability of reliable and accurate forecasts of the US tourism source market is of crucial importance, also in light of the results of recent surveys which show that the impact of the war in Ukraine could hurt travel sentiment to Europe in addition to COVID-19 health and safety concerns (European Travel Commission, 2022).
In the direction of tourism future research, we believe that our study contributes to the theme “tourism and technology” (Sharma et al., 2021) because it answers the research question on how e-tourism can be used for better markets forecast and furthermore raises awareness of foresight literacy in the tourism sector. Secondly, it complements the research on tourism demand forecasting by evaluating the performance of quantitative forecasting techniques on new data sources. The pandemic has reinforced the urgency to have new tourism data measures and complementary data sources to react quickly under uncertainty (OECD, 2022).

Thirdly, it is the first paper (to the best of our knowledge) that makes the United States arrival predictions in Italy. Despite its significance for destination and attraction management, the ability of the current approaches to precisely forecast visitor arrivals at particular attractions is still lacking (Volchek et al., 2019). Many works stop at the evaluation of the models, but few proceed in the forecast of the tourism demand (Huang and Yu, 2019; Li and Law, 2020; Antolini and Grassini, 2019). Finally, the findings of this study provide immediate valuable scenarios to private and public stakeholders in the tourism origin and destination markets that could be used to make decisions. The projected scenarios, derived with increased forecasting abilities, can help tourist managers choosing which management methods to consider and implement (Clark et al., 2019).

Bergman et al. (2010) distinguish different typologies of future studies according to making truth claims and/or indicating mechanisms: prediction, prognosis, science fiction and utopia (Bergman et al., 2010). However, a claim that is said to be true may be more or less conclusive. A forecast that assumes many outcomes, such as scenarios, based on which mechanisms are active is an example of this. Our work falls into this typology because we think it makes more sense to present a range of potential outcomes rather than a single-point forecast in the context of economic forecasting. The paper is structured as follows: Section 2 provides a brief background Section 3 presents Google Trends data collection and the statistical analysis. Finally, in Section 4 we discuss the result and draw some concluding remarks.

2. Background

Significant losses in service activity were brought on by COVID-19, particularly in the hospitality, food, entertainment and retail sectors and not just in the nations that strongly rely on tourists (Kanchana and Shyama, 2021). The pandemic period highlighted that tourism needs to be better measured, including its economic, environmental and social performance and it needs having of accurate, reliable and timely predictions. Achieving all this is certainly not easy due to the lack of historical data, seasonal fluctuations, influences of unexpected events like hurricanes and eclipses that affect visitation but are unrelated to the previous year’s visitation, the variety of input factors and the complexity of visitors’ travel decision-making process (Song et al., 2010; Clark et al., 2019). Certainly, these results can be achieved more easily favoring the digital transformation of the tourism industry (Akhtar et al., 2021) considered as information intensive (for instance with the implementation of tourism intelligence systems) and facilitating data access, fundamental elements of a big data-driven foresight approach. Foresight is defined as the process of foresight towards the future to try to find out what could be. This process appears significantly in future studies, where the foresight carries it out by applying the research methods based on orientation toward the future (Aldulaini et al., 2022).

As a result, demand modeling and prediction continues to receive a lot of attention from academics and industry professionals and are among the most important areas of tourist research (Hopken et al., 2019).

Several methods have been suggested to increase the forecasting precision of tourist demand: time-series analysis, econometric models, hybrid techniques, nonlinear modeling and artificial intelligence algorithms (Tian et al., 2021; Li and Law, 2020; Jun et al., 2018; Li et al., 2005). Among the quantitative methods applied to tourism demand forecasting transfer function models (describing the relationship between an output variable Yt at time t and k input variables X1t, . . . , Xkt) have also attracted the attention of scholars (Dinis et al., 2017; Gnanapragasam and Cooray, 2017).
Tourism demand often has a natural link with new web-based data sources like search engine traffic, website traffic or customer reviews on online review sites. These huge data sources have become more widely available, and as a result, forecasting tourist demand has also been done using them (Hopken et al., 2019). Literature has recently concentrated on taking advantage of the useful information search query data on tourist behavior provides. The number of travelers who use the internet to prepare for and book their business and leisure travels has dramatically increased over the past few years (Camacho and Pacce, 2018). Even for official statistics, big data are a vital source of information. International and national statistical offices are really conducting pilot studies on various big data sources today to see how well they would be able to provide or integrate official statistical data. Such data are primarily of importance because they can offer more time-saving and cost-effective information, especially at the local-level (Antolini and Grassini, 2019; Naccarato et al., 2018). Online data provide specific advantages over traditional data (e.g., survey-based data), such as higher frequency (i.e., daily, weekly, and monthly) and the capacity to reflect people’s behavior from different angles (Pan and Yang, 2017). By examining the themes that emerge from the online discourse and researching the dynamics of user interaction, big data from online social networks can be used to predict rapid changes in tourist preferences and popularity trends for destinations and local attractions (Franzetti Colladon et al., 2019). As a result, academics have added online data to traditional data, extracted business value from it and used it in the predicting models (Andariesta and Wasesa, 2022). The ability to track and reliably predict visitor behavior has been transformed by the availability of real-time, high-volume, and high-frequency data (Yang et al., 2015). Search data for tourist information have been extensively exploited to improve the forecasting capability of the current models (Park et al., 2017). A range of methods to apply search queries to predict tourist arrivals at a destination-level has been proposed (Antolini and Grassini, 2019; Artola et al., 2015; Bangwayo-Skeete and Skeete, 2015; Choi and Varian, 2009; Dergiades et al., 2018; Hopken et al., 2019; Li et al., 2018; Onder, 2017; Gunter and Onder, 2016; Park et al., 2017; Yang et al., 2015). The majority of existing studies demonstrate that online search engine data can improve the forecasting performance of tourism demand (Li and Law, 2020). Since
search query data reveals information about the interests, intentions and opinions of travelers, it is useful for predicting travel demand. Tourists utilize search engines to learn more about their vacation destinations, routes, must-see attractions and other travelers’ reviews and opinions of those locations. To predict the number of visitors who will actually visit a location, data from search queries may be employed.

Since ICTs and mobile computing devices have developed, the internet has become the primary source of travel information (Claude, 2020). Search engines, such as Google that dominate over a billion of worldwide Internet, Baidu and Yandex that represent the fastest growing emerging markets (i.e. China and Russia), shape tourist perceptions of attraction image and encourage travel decisions. In this paper, Google’s data were analyzed considering that most Americans have been using this search engine for some time to plan travel (Pan et al., 2011). Google began to release information on its users’ search queries through the publicly available Google Insights interface, now rebranded as Google Trends.

Since it began operating in 2004, Google Trends has provided search frequency of search keywords, which displays the comparison between the volume of searches for a given keyword and all other keywords over a given period of time. It then further normalizes search frequency into the range [0, 100], preventing changes in the volume of searches caused by an increase in users. The software displays the frequency of a specific keyword being used as a search term for the entire number of searches across multiple nations and languages (Choi and Varian, 2009). As a result, recent research has remarkably shown that Google Trends data reflect important elements of visitors’ keyword-based queries, which in turn opens up a wealth of options to research and predict travelers’ planned behavior (Dinis et al., 2019; Hopken et al., 2019; Yang et al., 2015; Kim et al., 2016; Hopken et al., 2021).

3. Google Trends data collection and analysis

3.1 Preliminary analysis

Considering the aim to predict international tourism demand in Italy using Google Trends (henceforth GT) data, first of all, we have proceeded to identify the possibly useful keywords. Following the literature suggestions (Camacho and Pacce, 2018; Havranek and Zeynalov, 2018; Hopken et al., 2019; Bangwayo-Skeete and Skeete, 2015; Claude, 2020) Holiday in Italy; Visit Italy; Travel to Italy; Hotels to Italy; Flights to Italy keywords have been initially selected. These expressions have been entered in GT to detect monthly search frequency data for each keyword from January 2015 to December 2021, carried out by Americans in the category Travel. After, we have computed the Pearson correlation coefficient between the GT keywords data and the target variable. Many different indicators have been used as the target variable in forecasting and modeling tourism demand. However, a commonly used measure is tourist arrivals. In particular, in this paper the target variable is represented by the Americans monthly arrivals in Italy from January 2015 to December 2020, carried out from the ISTAT website.

Correlation has been computed excluding the 2020 years, to avoid a possible bias due to the COVID-19 pandemic; however, unlike what one had been thought, correlations did not dramatically change also including 2020.

The highest correlation coefficients have been obtained for GT keywords Visit Italy (henceforth Visit) and Hotels to Italy (henceforth Hotel).

The choice of the GT keyword comparing the two candidates’ keywords has been made according to an intuitive criterion: we have chosen the GT keyword with a temporal dynamics which mimics as much as possible the temporal dynamics of the original series, using monthly data from 2015 to 2020. In practice, defined $X_t$ the i-th GT keyword at time t, we have compared the autocorrelation function (ACF) of $X_t$ to the ACF of the time-series Arrivals, denoted as $Y_t$. The ACF of the original time-series Arrivals is reported in Figure 1, while Figures 2 and 3 show the ACF’s of the two candidates.
The comparison is based on the quantity $\Delta \rho^M$, the average distance between $M$ ACF values of the series $X_t$ and $Y_t$:

$$\Delta \rho^M = \frac{\sum_{m=1}^{M} |\rho^X_m - \rho^Y_m|}{M}$$
where

\[ \rho_m^2 = \text{Corr}(Z_m, Z_{t-m}) \]

with

\[ Z \in (X, Y). \]

With \( \epsilon \in (13, 25, 37) \), the lowest value of \( \Delta \rho^M \) (see Table 1) is found for the GT keywords Hotel. The temporal dynamics of the time series Hotel is then as similar as possible to the original series Arrivals.

Figure 4 displays the time series of monthly arrivals up to December 2020. In Figures 5 and 6, the selected GT keyword is shown until 2020 and until 2021, respectively.

<table>
<thead>
<tr>
<th>GT keyword</th>
<th>M = 13</th>
<th>M = 25</th>
<th>M = 37</th>
</tr>
</thead>
<tbody>
<tr>
<td>Visit</td>
<td>0.269</td>
<td>0.235</td>
<td>0.209</td>
</tr>
<tr>
<td>Hotel</td>
<td>0.176</td>
<td>0.152</td>
<td>0.137</td>
</tr>
</tbody>
</table>

Source(s): Authors

Figure 4: Time plot of the monthly series Arrivals, 2015–2020

Source(s): Authors

Figure 5: Time plot of the monthly Google Trends keyword series Hotel, 2015–2020

Source(s): Authors
3.2 Transfer function

The use of GT data provides an insightful view of the event when original data are not available. However, practitioners and policymakers are not interested in the forecasting of GT data. The prediction of actual data is the primary goal. Defined $Y_t$ the time series of interest (Arrivals) and $X_t$ the observed GT keyword time series, we need to identify a transfer function (TF) from variable $X_t$ to variable $Y_t$,

$$TF : X_t \rightarrow Y_t$$

A general formulation is given by

$$\alpha(L)Y_t = \beta(L)X_t + v_t,$$

where

$$\alpha(L) = 1 - \alpha_1 L - \ldots - \alpha_p L^p$$

$$\beta(L) = 1 - \beta_1 L - \ldots - \beta_q L^q$$

Given the need of identifying a transfer function without using the lagged values of $Y_t$, we impose $\alpha(L) = 1$, so we focus on the expression:

$$Y_t = k + \beta(L)X_t + v_t$$

(1)

We identify the order $q$ that maximizes the adjusted coefficient of determination, that is $R^2$. The time period used in this step has to be covered by both the series, in particular the shortest one (Arrivals), that is January 2015–December 2020.

Table 2 reports the $R^2$ for Model 1 for $q = 0, 1, \ldots, 5$. The highest value is reached for $q = 4$ and the estimated parameters of the TF model are given in Table 3.

<table>
<thead>
<tr>
<th>Table 2</th>
<th>$R^2$ for TF model for $q = 0, 1, \ldots, 5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$q$</td>
<td>$R^2$</td>
</tr>
<tr>
<td>1</td>
<td>0.4764</td>
</tr>
<tr>
<td>2</td>
<td>0.6115</td>
</tr>
<tr>
<td>3</td>
<td>0.6538</td>
</tr>
<tr>
<td>4</td>
<td>0.6763</td>
</tr>
<tr>
<td>5</td>
<td>0.6741</td>
</tr>
</tbody>
</table>

Source(s): Authors
The TF equation is then
\[ Y_t = -429950.9 + 810.7X_t + 2927.2X_{t-1} + 4071.9X_{t-2} + 1170.4X_{t-3} + 3195.8X_{t-4} + \nu_t \] (2)

### 3.3 Statistical model for GT keyword time series

After estimating the TF, we have identified a model for GT keyword time series Hotel. The data now end at December 2021.

We have considered different models:

1. The first model is an ARIMA model without any intervention. The model naturally adjusts to the innovations. The resulting model is an ARIMA(0,1,0) \times (2,0,0)_{12} model,

\[
(1 - \Phi_1 L^{12} - \Phi_2 L^{24}) \Delta X_t = \epsilon_t
\]

and the estimates are reported in **Table 4**.

**Figure 7** reports the observed GT data and the fitted data from the estimated model.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Std.Error</th>
<th>t-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\kappa)</td>
<td>-429950.9</td>
<td>73086.3</td>
<td>-5.883</td>
<td>0.000</td>
</tr>
<tr>
<td>(\beta_0)</td>
<td>810.7</td>
<td>1396.2</td>
<td>0.581</td>
<td>0.564</td>
</tr>
<tr>
<td>(\beta_1)</td>
<td>2927.2</td>
<td>1956.0</td>
<td>1.497</td>
<td>0.140</td>
</tr>
<tr>
<td>(\beta_2)</td>
<td>4071.9</td>
<td>1981.7</td>
<td>2.055</td>
<td>0.044</td>
</tr>
<tr>
<td>(\beta_3)</td>
<td>1170.4</td>
<td>1937.7</td>
<td>0.604</td>
<td>0.548</td>
</tr>
<tr>
<td>(\beta_4)</td>
<td>3195.8</td>
<td>1428.1</td>
<td>2.238</td>
<td>0.029</td>
</tr>
</tbody>
</table>

**Source(s):** Authors

**Table 4** Estimate of the ARIMA(0,1,0) \times (2,0,0)_{12} model

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Std.Error</th>
<th>t-stat</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\Phi_1)</td>
<td>0.269</td>
<td>0.115</td>
<td>2.339</td>
<td>0.022</td>
</tr>
<tr>
<td>(\Phi_2)</td>
<td>0.348</td>
<td>0.142</td>
<td>2.451</td>
<td>0.016</td>
</tr>
<tr>
<td>(\sigma^2)</td>
<td>137.50</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>log-lik</td>
<td>-323.76</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AIC</td>
<td>653.53</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RMSE</td>
<td>11.515</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MAE</td>
<td>7.801</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Source(s):** Authors

**Figure 7** Time plot of the monthly Google Trends keyword series Hotel (black) and fitted values from the ARIMA(0,1,0) \times (2,0,0)_{12} model (red), 2015–2021

**Source(s):** Authors
2. The second model is an ARIMA model with an automatic procedure for the detection of possible transient level changes with a decay factor $|\delta| < 1$. The framework is now

$$X_t = X'_t + \sum_{i=1}^{\infty} f_i(t)$$

where $X_t$ is the observed (contaminated by the transient changes) series, $X'_t$ is the series we would observe in absence of the transient changes,

$$f_i(t) = \begin{cases} 
0 & \text{if } t < \tau_i \\
\omega_1 & \text{if } t = \tau_i \\
\omega_1 \delta_i & \text{if } t = \tau_i + 1 \\
\omega_1 \delta_i^2 & \text{if } t = \tau_i + 2 \\
\vdots & \vdots & \vdots \\
\vdots & \vdots & \vdots 
\end{cases}$$

With $\delta \in (0.70, 0.80, 0.90)$ assuming $\delta = \delta \forall i$, the model with the lowest RMSE is given by the ARIMA($1,0,0$) × (2,1,0)$_{12}$ with two transient changes ($M = 1$) at 63rd and 73rd times (i.e. March 2020–January 2021) and $\delta = 0.9$,

$$X_t = X'_t + f_1(t) + f_2(t)$$

where

$$f_1(t) = \begin{cases} 
0 & \text{if } t < 63 \\
-63.430 & \text{if } t = 63 \\
0.90(-63.430) & \text{if } t = 64 \\
0.90^2(-63.430) & \text{if } t = 65 \\
\vdots & \vdots & \vdots \\
\vdots & \vdots & \vdots 
\end{cases}$$

$$f_2(t) = \begin{cases} 
0 & \text{if } t < 73 \\
-36.378 & \text{if } t = 73 \\
0.90(-36.378) & \text{if } t = 74 \\
0.90^2(-36.378) & \text{if } t = 75 \\
\vdots & \vdots & \vdots \\
\vdots & \vdots & \vdots 
\end{cases}$$

The parameter estimates of the model are reported in Table 5.

Figure 8 shows the original (observed) series and the series after removing the estimated transient changes. In Figure 9 the observed GT series is plotted together with the fitted series from the estimated model.

The higher flexibility of model allowing for transient changes involves a better fit, as confirmed by RMSE and MAE, which are 11.515 and 7.801 for the former model and 6.630 and 4.927 for the latter.
3.4 Predictions

Using the TF estimated in subsection 3.2, we can predict the original series Arrivals using a two-step procedure. First, we predict the GT keyword Hotel for 2022. The point forecasts as well as the interval forecasts with a confidence level of 95% are shown in Figure 10.
Applying equation (2), we obtain for the year 2022 the mean prediction when the input is given by the point forecast (Figure 11). This is the expected neutral scenario in absence of new relevant occurrences. However, the possible rebound of the tourist activity after two difficult years suggests considering an optimistic scenario as well. In this case, the input of the TF is represented by the upper limit of the interval forecast at the 95% confidence level, that is, the 90th percentile. The higher-level predictions are presented in Figure 12.

4. Conclusions

This work adopts a big data-driven foresight approach to predict the international visitor arrivals in Italy using Transfer Function on the US Google Trends searches data. Our analysis remembering the emerged tech–savvy generations in the evolving visitor demand megatrend demonstrate Google Trends’ excellent potential for forecasting tourism demand in timely manner, particularly useful in a situation where events, like the COVID-19 outbreak or the Russian–Ukrainian conflict, have increased uncertainty in travel planning. Airline tickets are being booked increasingly close to departure, with a reduction between purchase and actual passage, for example, on average for Americans it has gone from 103 to 80 days, while for Japanese even from 76 to 32 days. Keeping in mind that the theory of forecasting is based on the premise that the current and past knowledge can be used to make predictions about the future (Petropoulos et al., 2022), the results obtained...
from complementary between theoretical developments and applications using transfer function on Google Trends data are in line with the opportunity to implement foresight approach based on new instruments in the tourism industry.

A digital transformation based on digital infrastructure, data strategy, augmented/virtual reality would allow the tourism industry to benefit from timely forecasts for smart and better decision-making.

In this paper, two different models using monthly data have been evaluated. The first model is an ARIMA(0,1,0) x (2,0,0)_{12} model without any intervention. The second model allows transient-level changes (in particular a transient level change is expected in March 2020). An automatic procedure for the detection of possible transient level changes has been applied. Two scenarios emerged from the analysis characterized by different specific conditions. The first scenario, could be defined as prudent where the effect of perceived potential risk on travel and tourism decision-making after the COVID-19 crisis are still present (Matiza, 2022) and where forecasts are still struggling to reach pre-pandemic levels and the second, defined as optimistic in which the desire to be able to return to travel means that the average forecasts are in fact exceeded by the reality. The second scenario envisaged is reflected both in the data of the first half of 2022 which see an turn of tourist flows in Italy, including international ones and among these, above all Americans (also thanks to a favorable euro/dollar exchange rate) and in the greater confidence in the future on the part of tour operators.

If the hypothesis of prudent scenario is compatible with niche phenomena such as tourism zero kilometer and slow (Monaco, 2021), with a resilient approach to adaptation and transformation (Kwok and Koh, 2022) and with the promotion of domestic tourism to fill the gaps left by international tourism, the optimistic one need evaluation of possible over-tourism phenomena, of safety in tourism (Monaco, 2021) and in general, of tourism well-being above all in a sustainability perspective (Larry, 2020).

While aware that factors such as subjectivity in the choice of keywords, Google Trends data quality and limitations of transfer function used only for linear systems (Lai, 1979) could have influenced the results obtained, it is believed that the path identified can make a significant contribution to take proactive measures to reduce uncertainty and effectively prepare for megatrends that will significantly impact tourism. In a future perspective, it could be interesting to observe the scenario generated by the application of the methodology to other origin markets of the international tourism in Italy or by data carried out on other search engines. Furthermore, external variables could be included in the analysis, considering a transfer function in which some external variables are added to the Google Trends values. In addition, obtaining forecasts through the use of transfer function models on Google Trends data could be a valid methodological tool also for other socioeconomic or environmental phenomena. For example, stakeholders of new emerged forms of tourism (i.e.

**Figure 12** Optimistic predictions of arrivals for 2022 based on the 90 percentile of the predictor: real data in black and predictions in red

![Optimistic predictions of arrivals for 2022 based on the 90 percentile of the predictor: real data in black and predictions in red](https://via.placeholder.com/150)
wellness, health, volunteering and bleisure) could benefit from the forecasts obtained with the transfer function models from Google searches. Moreover, it could be useful to mitigate the vulnerability of the tourism sector due to climate change (Leal Filho, 2022).

In conclusions, the availability of a plurality of scenarios allows stakeholders to evaluate in a more coherent way the implications of political choices that can range from investment to restore consumer confidence, to seize the marketing opportunities offered by virtual reality (Mohanty et al., 2020; Monaco and Sacchi, 2023), to improve digital skills and infrastructure (Knezevic et al., 2021) to produce quick response and effective decision-making for future crises, to use of targeted and personalized marketing channels and quality improvement.

References


About the authors
Giovanni De Luca is Professor of Economic Statistics at the Department of Management and Quantitative Studies, University of Napoli “Parthenope”. He teaches time-series analysis. His research interests include time-series analysis, forecasting and multivariate analysis.

Monica Rosciano is Assistant Professor of Economic Statistics at the Department of Management and Quantitative Studies, University of Napoli “Parthenope”. She teaches statistical models for data analysis. Her research concerns the issues of financial well-being, tourism competitiveness and sustainability. Monica Rosciano is the corresponding author and can be contacted at: monica.rosciano@uniparthenope.it

For instructions on how to order reprints of this article, please visit our website: www.emeraldgrouppublishing.com/licensing/reprints.htm
Or contact us for further details: permissions@emeraldinsight.com