Journal of Tourism Management Research
2024 Vol. 11, No. 2, pp. 302-312
ISSN(e): 2313-4178
ISSN(p): 2408-9117
DOI: 10.18488/31.v11i2.3986

Is it possible for Google Trends to forecast rural tourism? The situation in Spain

Eduardo Cebrian[1+]
Josep Domenech[2]

[1,2]*Department of Economics and Social Sciences, Universitat Politècnica de València, Spain.*
[1]*Email: ecebcer@esp.upv.es*
[2]*Email: jdomenech@upvnet.upv.es*

(+ Corresponding author)

Article History
Received: 3 September 2024
Revised: 28 October 2024
Accepted: 11 November 2024
Published: 22 November 2024

Keywords
Forecasting
Google Trends
Rural development
Rural tourism
Tourism
Tourism demand.

ABSTRACT

The study aims to assess the potential of Google Trends (GT) data for improving the prediction of monthly rural overnight stays by national residents in Spain. The study uses forecasting models that incorporate Google Trends data and compares their accuracy with traditional time-series benchmark models. The data used spans from January 2012 to February 2020. This comparison allows for a direct evaluation of whether GT variables offer any predictive advantage over classical time-series methods in the context of rural tourism. The results indicate that models incorporating GT variables do not outperform traditional time-series models in predicting rural tourism flows in Spain. Previous studies in other tourism sectors have found that GT data enhances predictive accuracy. The study concludes that Google Trends data may have limitations in predicting rural tourism flows, despite its demonstrated utility in other tourism forecasting contexts. These limitations suggest a need for further investigation into why GT does not improve rural tourism forecasts. The findings suggest that tourism industry stakeholders and policymakers should be cautious in relying on GT data for forecasting rural tourism flows.

Contribution/Originality: In this study, we employ Google Trends data (GT) to test whether using GT data can improve forecasting results for rural touristic flows in Spain. Forecasting regular destinations has proven the success of this approach, but forecasting rural flows presents a novel challenge.

1. INTRODUCTION

In recent decades, the rise of the Internet and online booking platforms has revolutionized rural tourism, making it more accessible and attractive than ever. By 2021, rural areas accounted for a striking 43.8% of accommodation beds and contributed to 37% of all overnight stays across the European Union (EPRS, 2023).

Spain, where tourism plays a vital role in the economy, saw the sector contribute 12.4% of its GDP in 2019 (UNWTO, 2023). Moreover, from 2012 to 2019, rural tourism in Spain surged by over 20%, driven by national residents seeking countryside escapes (INE, 2023).

As rural tourism is comprised of activities that take place in areas with low population density and the landscape is tied to agriculture and nature, it can be a powerful driver for rural development by fostering sustainable local employment and boosting profitability in these regions (Guaita Martínez, Martín Martín, Salinas Fernández, & Mogorrón-Guerrero, 2019; Wijijayanti, Agustina, Winarno, Istanti, & Dharma, 2020).

For this reason, the scientific literature has focused on identifying the key drivers that promote this development (Kumar & Valeri, 2022; Liu, Dou, Li, & Cai, 2020; Rid, Ezeuduji, & Pröbstl-Haider, 2014; Yang et al.,

2021) while also addressing the challenges that may hinder this growth process (Rosalina, Dupre, & Wang, 2021; Wang et al., 2013).

A key part of developing touristic areas is to accurately predict their touristic flows. Improved forecasts of touristic demand can aid management as well as policy makers in these rural areas in making decisions about the planning and strategies in tourism (Bangwayo-Skeete & Skeete, 2015; Song, Witt, & Li, 2008), as they can provide vital information in aspects such as the definition of price policies or allocation of resources (Colladon, Guardabascio, & Innarella, 2019; Dergiades, Mavragani, & Pan, 2018) and, in turn, facilitate the development of these rural areas.

In tourism research, one of the ways of improving touristic forecasts is to incorporate data from Google Trends (GT) as a predictor to enhance the accuracy of tourist flow forecasts (Bangwayo-Skeete & Skeete, 2015; Havranek & Zeynalov, 2021; Rivera, 2016). This method involves comparing the forecasting performance of models that utilize GT data with those that do not. While this approach has proven effective in traditional tourism contexts (Bokelmann & Lessmann, 2019; Park, Lee, & Song, 2017; Wickramasinghe & Ratnasiri, 2021), its application in rural tourism remains underexplored. As a result, the potential for GT data to enhance forecasting accuracy in rural tourism contexts is yet to be determined, highlighting a gap in the literature that merits further investigation due to its potential benefits.

This paper fills in the gaps in the research by checking how well models that use GT data and traditional time-series benchmark models can predict overnight stays for tourists in rural Spain from January 2012 and February 2020. The objective is to assess whether the inclusion of GT data enhances the accuracy of these forecasts.

The rest of this paper is structured as follows: Section 2 reviews relevant literature, demonstrating the potential benefits of rural tourism and the capacity of GT to improve forecasting of variables across different scientific fields. Section 3 describes the methodology for the extraction and treatment of the data as well as the chosen autoregressive models. Section 4 provides the results regarding, first, the fit of the competing models and second, the forecasting performance of the different models. Finally, Section 5 provides some discussion as well as some concluding remarks.

2. LITERATURE REVIEW

UNWTO (2019) defines rural tourism as *'a type of tourism activity in which the visitor's experience is related to a wide range of products generally linked to nature-based activities, agriculture, rural lifestyle/culture, angling, and sightseeing'*. Therefore, rural tourism is a collection of activities that take place in destinations with low population density, where the landscape is predominantly shaped by agriculture and forestry, and where there is a predominantly traditional lifestyle. In fact, Rosalina et al. (2021) argue the analysis of other definitions for rural tourism reveals four key features: rural tourism can be defined by its location, by its efforts towards sustainable development, by its community-based features, and by the kind of experiences it provides.

Regarding location, it refers to rural tourism occurring in rural areas, even if some authors argue that this does not always have to be true (Komppula, 2014). Yet, authors mostly employ rural areas as case studies. For instance, Jepson and Sharpley (2015) conduct interviews with visitors to the Lake District area (UK) and discover that visitors feel a sense of belonging to the area, which is developed through its scenery and through participation in certain forms of rural tourism that evoke deeper emotional connections, suggesting that creating 'a sense of place' plays a significant role for tourists. In a similar fashion, Rocca and Zielinski (2022) use Minca, Sierra Nevada de Santa Marta (Colombia) as a case study to explore the relationships between community-based tourism, social capital, and governance in post-conflict contexts.

Secondly, sustainable development focuses on fostering the growth of rural tourism in a sustainable manner. In this context, Whitelaw, King, and Tolkach (2014) present different proposals to fund and manage protected areas in a sustainable way, as they are becoming a popular spot for tourists. Rosalina, Dupre, Wang, Putra, and Jin (2023) propose different management strategies of touristic resources for two Balinese villages that allow for resource conservation and an inclusive local economy while also including the Balinese culture and spiritual component.

Thirdly, community-based features refer to the participation and coordination of different economic agents, such as the local communities and the governments. Liu et al. (2020) explain the role played by both the national and the local governments in the stimulation policies used to boost rural regeneration in China. The central government creates an environment where local governments are able to manage tourism and provide services in a more direct manner. Similarly, Yang et al. (2021) study the effects of rural revitalization, a strategy used to boost rural economies where the government provides political measures and financial support, and later on, businesses use these resources to foster rural development. They find an improvement in both the average building height and the proportion of non-agricultural employment from 1988 to 2016 in the region of Liaodong, China.

Finally, experiences refer to the physical, social, and psychological experiences of tourists. In this case, experiences tend to be linked to certain activities that can evoke positive feelings in the customer, such as wine tourism through its history and brand (Bonarou, Tsartas, & Sarantakou, 2019; Frost, Frost, Strickland, & Maguire, 2020) or by using the features of the natural terrain as the basis for an adventure activity (Buckley, 2007). Moreover, these feelings can be caused as well through sensory experiences that cause an attachment to the destination (Kastenholz, Marques, & Carneiro, 2020).

Nevertheless, one of tourism's greatest benefits is its enormous potential to boost rural economies by offering a variety of revenue streams and strengthening the socioeconomic fabric of rural areas (Martínez, Martín, Fernández, & Mogorrón-Guerrero, 2019; Wijijayanti et al., 2020). Numerous factors contribute to rural tourism's economic impact, such as promoting economic growth, reducing poverty, and raising living standards in local communities (Liu, Chiang, & Ko, 2023; Wilson, Fesenmaier, Fesenmaier, & Van Es, 2001). Moreover, the clustering of tourism-related activities tends to promote cooperation and partnerships among local actors, leading to such growth (Kumar & Valeri, 2022; Wilson et al., 2001). By guaranteeing the sustainability of these tourism ventures, this cooperative framework not only strengthens local economies but also builds a strong economic environment (Nooripoor, Khosrowjerdi, Rastegari, Sharifi, & Bijani, 2021).

Beyond economic growth, tourism makes a substantial contribution to rural communities' sociocultural and environmental well-being. It improves the standard of living for locals and offers opportunities for cultural preservation, social stability, and community pride (Liu et al., 2023). The sociocultural advantages are also manifold, as they encompass the revitalization of regional traditions and crafts as well as the advancement of cultural heritage and communal identity (Lane & Kastenholz, 2015). In terms of the environment, tourism can result in stronger conservation initiatives that support biodiversity and sustainable land use methods. The transition towards tourism makes it possible for rural communities to preserve their scenic and natural beauty, attracting tourists and ensuring long-term survival (Jepson & Sharpley, 2015; Lane & Kastenholz, 2015; Rosalina et al., 2021). Because it addresses economic, sociocultural, and environmental aspects of rural development and promotes a sustainable future for rural communities, the incorporation of tourism into rural economies offers a comprehensive approach to rural development.

One of the most important aspects of developing and planning tourism is accurately forecasting visitor flows, as it allows economic agents to make informed decisions, which may lead to the development of the area. In this regard, there is a wealth of scientific literature that studies the forecasting of touristic flows. In terms of methods, time series and econometric techniques seem to be most popular (Jun, Yoo, & Choi, 2018; Li, Law, Xie, & Wang, 2021). More specifically, researchers tend to employ different model specifications and compare their forecasting performance. Most authors tend to use some specification of autoregressive models (AR) (Bangwayo-Skeete & Skeete, 2015; Li, Pan, Law, & Huang, 2017) or Vector Autoregressive Models (VAR) (Dergiades et al., 2018; Liu, Tseng, & Tseng, 2018) as one of the competing models in the forecasting while also including some benchmark models such as a Holt Winters (HW) or Naïve specification (Önder, 2017; Rivera, 2016).

In terms of data sources, search engine data are the most common to improve forecasting performance, and within them, GT seems to be the most popular. Google developed GT as a search engine data tool. Part of its popularity among the scientific community may be due to the fact that its data is publicly available, frequently updated, and its interface is simple to use (Jun, Park, & Yeom, 2014). GT provides an index of the relative popularity of queries performed on Google (Cebrián & Domenech, 2023). This data is available from 2004 to the present day. The index is based on a query share, where the point in time with the highest number of searches for the chosen search term is set to 100 and the remaining points are calculated in relative terms from it Choi and Varian (2012). You can limit index to a specific time and geographic area. Additionally, searches are categorized into groups like 'Food & Drinks' (category 71) or 'Health' (category 45). These groups serves as filters, presenting only searches associated with the category that are chosen in conjunction with a search term.

In this vein, Park et al. (2017) employ models that include GT searches to improve forecasts of Japanese inflows to Korea, while Bangwayo-Skeete and Skeete (2015) successfully improve forecasts of hotel and flight data from the US, Canada, and the UK to five Caribbean destinations. However, not all authors consistently find that GT improves tourism forecasting. For instance, Rivera (2016) uses GT data to enhance forecasting predictions of the number of non-resident registrations in Puerto Rican hotels and finds that Holt-Winters models outperform models with GT data in short-term forecast horizons, but linear models with GT data tend to outperform in longer forecast horizons.

However, the forecasting of touristic flows in the context of rural tourism is not as well explored in the literature. For instance, Yin (2020) effectively adjusts forecasting techniques to recently developed rural China areas, while Xi and Donglai (2022) forecast regional flows of rural tourism to Jiayuguan, China, using an improved Quad-Res Net model and achieve better forecasting outcomes.

Despite these results, both authors draw attention to the possible challenges that may arise from rural tourist flows, noting the possibility of seasonal effects and high levels of volatility (Xi & Donglai, 2022). Moreover, the absence of adequate historical market data and the limited online presence of newly established rural areas pose significant challenges. Yin (2020), which could thereby reduce the impact of conventional techniques for predicting touristic flows.

We identify a literature gap since, to the best of our knowledge, no research has been done on the use of GT data as a predictor for rural tourism flows. Do these established tourist techniques also apply to rural tourism? Can we forecast inflows of rural tourists in the same manner that we forecast inflows of overall tourists? Consequently, this article's contribution is the forecasting of national residents' rural overnight stays in Spain using Autoregressive Distributed Lag models (ADL), which include information from GT, and determining whether or not these models outperform benchmark time-series models.

3. METHODOLOGY

The methodology section is divided into two parts: First, Subsection 3.1 describes the data employed in the manuscript, pertaining to both the rural overnight stays as well as that obtained from GT. Secondly, Subsection 3.2 details the required treatments performed on the data. Finally, Subsection 3.3 explains the techniques utilized to obtain the different model specifications.

3.1. Data

3.1.1. Rural Overnight Stays

The dependent variable in this study is the monthly rural overnight stays by national citizens in Spain from January 2012 (2012M1) to February 2020 (2020M2). In order to prevent the COVID-19 pandemic from introducing a shock into the dataset and maybe making cross-model comparisons more difficult, this time frame is specifically chosen. The information is taken from a survey that the INE (Instituto Nacional de Estadística), Spain's official statistical agency, conducted on the occupation of rural tourism accommodations (officially called 'Encuesta de ocupación en alojamientos de turismo rural').

3.1.2. Google Trends data

The first step in using data from GT is to choose keywords based on search queries that may be associated with the dependent variable. In order to accomplish this, a set of various suggested combinations of search terms and categories is extracted, and the correlation between each combination and the dependent variable is then calculated. Using R-Studio and the 'trendecon' R-package (Eichenauer, Indergand, Martínez, & Sax, 2022), the extraction is performed, and the data is obtained at a monthly frequency that corresponds to the dependent variable. The combinations with the highest Pearson correlation coefficient are then selected for further action from among those that have significant correlations to the dependent variable Table 1. The suggested search terms are 'agroturismo' (agrotourism), 'turismo rural' (rural tourism), and 'rural" (rural), and the suggested categories are 67 'travel', 1389 'agrotourism', 1005 'ecotourism,' and 1391 'vineyards and wine tourism'. All of them are retrieved for the geographic area of Spain, and ultimately, the following combinations are selected: 'rural' under the 'travel' and 'agrotourism' categories, and 'agroturismo' under the same categories.

Table 1. Search term and category combinations and their correlation tothe dependent variable.

Search term	Category	Correlation
Rural	**Travel**	**0.50*****
Rural	**Agrotourism**	**0.66*****
Rural	Ecotourism	0.37***
Rural	Vineyards & wine tourism	0.04
Turismo rural	Travel	0.35***
Turismo rural	Agrotourism	0.29***
Turismo rural	Ecotourism	0.17**
Turismo rural	Vineyards & wine tourism	-0.01
Agroturismo	**Travel**	**0.71*****
Agroturismo	**Agrotourism**	**0.69*****
Agroturismo	Ecotourism	0.05
Agroturismo	Vineyards & wine tourism	-0.01

Note: Rows in bold indicate the selected queries for the next steps in the study.
***Statistical significance at the 1% level, **Statistical significance at the 5%level.

Table 2. Necessary extractions to obtain a 1% MAPE for each search term and category combination.

Search term	Category	S.D. (10 extractions)	Extractions	Extractions rounded
Rural	Travel	1.72	5.64	6
Rural	Agrotourism	2.50	12.20	13
Agroturismo	Travel	2.75	14.74	15
Agroturismo	Agrotourism	3.76	28.55	29

Next, the Cebrián and Domenech (2024) approach for GT data processing is employed, given that different extractions can generate non-negligible difficulties with data accuracy as they fluctuate day to day (Cebrián & Domenech, 2023). The first step is to find the mean standard deviation (s.d.) of the various extractions. To do so, the chosen combinations are first extracted ten times each. Then, the formula is applied to the mean s.d. of each combination to determine the required number of extractions to get at a 1% Mean Absolute Percentage Error (MAPE) for each of the chosen combinations. The results of this process are presented in Table 2.

3.2. Data Treatment

Following their extraction, the search terms are converted into logarithms together with the rest of the data.

Then, a multiplicative method is used to deseasonalize the data. Next, unit root tests are run on all the variables, and it is found that there is only one unit root for the various GT terms as well as for rural overnight stays, and for that reason, all ofthe variables are used in first differences.

3.3. Models

Then, a number of Autoregressive Distributed Lag Models (ADLMs) are fitted. These include a baseline model that uses the dependent variable's 12 lagged values as the only predictor and a set of models that extend the baseline specification by including the 12 lagged values of the various GT search terms as well as the contemporary value of the GT search term.

Furthermore, in order to account for both linear and non-linear trending behavior, each model contains a quadratic trend component in addition to a linear trend component.

Lastly, using the entire sample (2012M1-2020M2), the models are estimated using Ordinary Least Squares (OLS) with robust heteroskedasticity errors. Equation 1 provides a general specification of the models, where the logarithm of overnight stays is denoted by Y_t, the lags of the dependent variable are represented by Y_{t-i}, the lags for the GT term are denoted by GT_{t-i}, and the linear and quadratic trend components are denoted by t and t^2, respectively[1].

$$Y_t = \alpha + \sum_{i=1}^{12} \gamma * Y_{t-i} + \sum_{i=0}^{12} \beta * GT_{t-i} + \delta * t + \zeta * t^2 + \epsilon_t \quad (1)$$

4. RESULTS

The Results Section is divided into two parts: First, Subsection 4.1 describes the fit of the models specified, and secondly, Subsection 4.2 dives into the forecasting performance of said models and compares itsresults.

4.1. Estimation Results

The econometric models are estimated as specified in Subsection 3.3, and the results for the in-sample estimations of all models are presented in Table 3.

First, all the models present high overall significance, as shown by their F-statistics, which provide significant values for all the models. Then, once all the models can be considered valid, comparison among them becomes relevant. In this case, Model 4 provides the best fit across all models with an adjusted R^2 of 0.563, and Model 5 presents the lowest fit with an adjusted R^2 of 0.488.

There are also notable differences in the significance of the lags of the dependent variable across models. While the first four lags of the dependent variable are significant in all the models, then the significance of the rest of lags varies across models. For example, in Models 2 and 3, a majority of the rest of lags of the dependent variable are significant, while in Model 4, only 5 lags are significant in total.

There are a few other significant distinctions with respect to the GT terms. First off, Model 5 is the least well-fitting model since it lacks a significant GT term, lagged or not. The three best-fitting models, which are 2, 3, and 4, are so by a decent margin, and each of them offers at least two significant GT terms. Specifically in Model 2, out of the 13 GT terms, 10 are significant at 10%.

Finally, none of the models show any significant trending pattern, linear or otherwise.

[1] This study's technique is based on an adaption of Önder (2017).

Table 3. In-sample fit model estimations.

Model	Model (1)	Model (2)	Model (3)	Model (4)	Model (5)
Search term category	- -	Rural travel	Rural agrotourism	Agrotourism travel	Agrotourism agrotourism
Constant	0.010	0.022	0.012	0.011	0.013
y (-1)	-0.808***	-0.827***	-0.746***	-0.766***	-0.809***
y (-2)	-0.659***	-0.737***	-0.641***	-0.656***	-0.676***
y (-3)	-0.599***	-0.778***	-0.587***	-0.653***	-0.677***
y (-4)	-0.518**	-0.736***	-0.579***	-0.477**	-0.544***
y (-5)	-0.409*	-0.727**	-0.440**	-0.307	-0.409**
y (-6)	-0.225	-0.506**	-0.239	-0.098	-0.210
y (-7)	-0.273	-0.594***	-0.305*	-0.039	-0.250
y (-8)	-0.291	-0.558***	-0.317*	-0.053	-0.240
y (-9)	-0.363**	-0.591***	-0.390**	-0.193	-0.340**
y (-10)	-0.352**	-0.528***	-0.312**	-0.110	-0.310*
y (-11)	-0.104	-0.216	0.077	-0.025	-0.133
y (-12)	-0.300**	-0.409***	-0.367***	-0.280**	-0.305***
GT	-	0.155	0.071**	0.058*	0.024
GT (-1)	-	0.228**	0.085*	0.082**	0.051
GT (-2)	-	0.234**	0.072	0.079	0.067
GT (-3)	-	0.294***	-0.012	0.098*	0.070
GT (-4)	-	0.196*	0.064	00.062	0.064
GT (-5)	-	0.330**	0.046	0.040	0.037
GT (-6)	-	0.204*	0.022	0.025	0.021
GT (-7)	-	0.258**	0.002	0.038	-0.008
GT (-8)	-	0.198	0.006	-0.022	-0.027
GT (-9)	-	0.302***	0.044	0.001	0.016
GT (-10)	-	0.211*	0.005	-0.023	-0.030
GT (-11)	-	0.076	-0.031	-0.021	-0.008
GT (-12)	-	0.191*	0.008	0.053*	0.021
T	0.000	0.000	0.000	0.000	0.000
T^2	-0.000	-0.000	-0.000	-0.000	-0.000
Adjusted R^2	0.501	0.553	0.548	0.563	0.488
F-statistic	8.265	7.926	8.751	8.795	5.897
P-value	$4.308e^{-10}$	$3.549e^{-11}$	$4.741e^{-12}$	$4.278e^{-12}$	$9.654e^{-9}$

Note: ***Statistical Significance at the 1% level, **Statistical Significance at the 5% level, *Statistical Significance at the 10% level.

4.2. Forecasting Results

In this section, the out-of-sample forecasting performance of the models is compared through the Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE). Additionally, in line with Önder (2017), rival time series models are included, including a naïve model and a non-seasonal Holt-Winters model to the baseline model and the GT augmented models (Models 2 through 5).

Next, a rolling window methodology is used to determine the performance of the models at all forecasting horizons. Then, Table 4 displaysthe process's outcomes for predicting horizons of 1, 2, 3, 6, and 12.

For all forecast horizons, the Naïve model is the best performing model, followed by the Holt-Winters model, with the baseline model beingthird best.

Lastly, out of the models that contain GT terms, Model 3 offers the greatest forecasts, while Model 5 offers the poorest. However, even Model 3 is not able to outperform the Naïve, Holt-Winters, or baseline models at any point.

Moreover, spiking behavior is found as well regarding the RMSE and MAE in the baseline and the GT models. In this case, the spiking behavior occurs in horizons h=1 and h=2, where both the RMSE and MAE values are higher than they are at the rest of the forecasting horizons. From h=3 onward, both statistics start at a lower value and then steadily increase, which is a more typical behavior. Since their model fits are not very high to begin with, the spikes shown at the first two horizons might be the result of a substantial amount of noise in the prediction.

Overall, it cannot be said that GT data improves predictions for rural overnight stays by national residents in Spain, at least not with this specification, since the models with a GT predictor do not outperform classical time series models like a Naïve or a non-seasonal Holt-Winters model.

Table 4. Out of sample forecasting results, forecast horizons 1, 2, 3, 6 and 12.

h = 1			h = 2		h = 3		h = 6		h = 12	
Model	**RMSE**	**MAE**	**RMSE**	**MAE**	**RMSE**	**MAE**	**RMSE**	**MAE**	**RMSE**	**MAE**
Model (2)	0.273	0.109	0.301	0.114	0.153	0.094	0.130	0.097	0.165	0.127
Model (3)	0.097	0.075	0.098	0.074	0.089	0.068	0.098	0.072	0.128	0.094
Model (4)	0.129	0.074	0.146	0.079	0.129	0.075	0.201	0.095	0.308	0.113
Model (5)	0.926	0.205	0.903	0.193	0.374	0.117	0.105	0.076	0.243	0.119
Baseline	0.075	0.052	0.081	0.056	0.083	0.058	0.085	0.065	0.114	0.082
Naïve	0.046	0.032	0.047	0.032	0.042	0.030	0.042	0.029	0.044	0.030
Holt-Winters	0.052	0.043	0.054	0.045	0.056	0.047	0.056	0.049	0.066	0.060

Yet, these findings are in line with those of Önder (2017) who also find that the Naïve specification outperforms all other competing models in terms of arrival prediction for Belgium when total searches, searches from the U.S., and searches from the UK are employed as predictors. In a similar vein, Gunter, Önder, and Gindl (2019) also discover that Naïve models perform better than models that solely use lags of the dependent variable and lags of GT as the primary predictors when attempting to forecast the total number of tourists that will visit four Austrian cities. Further evidence is found by Rivera (2016) indicating that the Dynamic Linear Model (DLM) employed to estimate non-resident registrations for Puerto Rico hotels only beats other benchmark models like the ones in this study when the forecasting horizon is six or greater.

5. CONCLUSIONS

5.1. Main Implications

The acquired results do not appear to support the preliminary hypothesis that using GT data enhances the forecasts of monthly rural overnight stays by Spanish nationals. As demonstrated in Section 4.2, all the models that do not include GT information outperform those that do. Since not all users use Google search engines to find information on their rural tourism activities, one explanation for these results could be that GT does not capture a sufficient amount of the search information produced by tourists. Another reason might be that, if tourists indeed use Google search engines, they might search for the specific places that they might want to visit rather than searching for more generalist queries such as the ones suggested in the study. As a result, queries that include more general terms, such as 'rural tourism,' might not be able to capture the touristic behavior as intended.

This is illustrated in Subsection 4.1, where it is shown that models using GT queries do not yield a significantly better fit compared to those that do not. Furthermore, even the models that do offer a better fit (Models 2, 3, and 4) present adjusted R^2 lower than 0.6, indicating that over 40% of the dependent variable's variability is not accurately captured by this specification, even if these specifications have yielded positive results for other authors in the past (Önder, 2017), albeit in different contexts.

However, even if tourists search for specific locations as keywords, it would be very difficult to generate an estimate of the searches for rural tourism for a number of reasons. The first of these reasons is the increased expense of finding every location and search term associated with the dependent variable. Secondly, consistent estimations of the searches conducted for each location would be very challenging to obtain, given that GT does not do well with low popularity queries (Cebrián & Domenech, 2024). Furthermore, GT may report zeros if the search volume falls below a specific popularity threshold (Cebrián & Domenech, 2023), which could be the case for less popular rural locations, and in this scenario, it would be impossible to create a GT index at all. Thirdly, it would not be possible to create an aggregated index either, even if the searches could be retrieved for every query, as GT does not offer the raw data of these searches. Additionally, it is likely that some information is shared through channels other than the internet, such as word-of-mouth or other channels that search engines are unable to account for.

Lastly, it is also possible that for any of the previously mentioned reasons or for other reasons that may escape our knowledge, GT does not predict rural tourism variables as well as it does other types of tourism or other sorts of variables.

5.2. Limitations

This work also faces some limitations, which are listed next: Initially, to try to improve the forecasting results, several search terms could have been combined instead of combining search terms and categories. Second, this study confines its findings to the scenario of national citizens spending the night in rural Spain for tourism purposes. As a result, different circumstances or econometric specifications mayyield different findings.

5.3. Future Research

Lastly, future research should examine these results in various settings to confirm whether GT can be helpful in enhancing forecasts for rural tourism. More specifically, future research should focus on finding different combinations of search terms or search terms and categories that can help corroborate the findings in this manuscript. Moreover, future research should also attempt to reproduce this analysis in a wider variety of markets and countries and through different model specifications.

Funding: This work was partially supported by grant PID2023-152106OB- I00 funded by MCIN/AEI/10.13039/501100011033 and by ERDF/EU and by grant CIAICO/2023/272 funded by Generalitat Valenciana.
Institutional Review Board Statement: Not applicable.
Transparency: The authors state that the manuscript is honest, truthful, and transparent, that no key aspects of the investigation have been omitted, and that any differences from the study as planned have been clarified. This study followed all writing ethics.
Competing Interests: The authors declare that they have no competing interests.
Authors' Contributions: Original writing of the manuscript, analysis and methodology, E.C.; conceptualization, supervision and review of the manuscript, J.D. Both authors have read and agreed to the published version of the manuscript.

REFERENCES

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. https://doi.org/10.1016/j.tourman.2014.07.014

Bokelmann, B., & Lessmann, S. (2019). Spurious patterns in Google trends data-An analysis of the effects on tourism demand forecasting in Germany. *Tourism Management, 75*, 1-12. https://doi.org/10.1016/j.tourman.2019.04.015

Bonarou, C., Tsartas, P., & Sarantakou, E. (2019). *E-storytelling and wine tourism branding: Insights from the "wine roads of Northern Greece*. Cham: Springer International Publishing.

Buckley, R. (2007). Adventure tourism products: Price, duration, size, skill, remoteness. *Tourism Management, 28*(6), 1428-1433. https://doi.org/10.1016/j.tourman.2006.12.003

Cebrián, E., & Domenech, J. (2023). Is Google trends a quality data source? *Applied Economics Letters, 30*(6), 811-815. https://doi.org/10.1080/13504851.2021.2023088

Cebrián, E., & Domenech, J. (2024). Addressing Google trends inconsistencies. *Technological Forecasting and Social Change, 202*, 123318. https://doi.org/10.1016/j.techfore.2024.123318

Choi, H., & Varian, H. (2012). Predicting the present with Google trends. *Economic Record, 88*(1), 2–9. https://doi.org/10.1111/j.1475-4932.2012.00809.x

Colladon, A. F., Guardabascio, B., & Innarella, R. (2019). Using social network and semantic analysis to analyze online travel forums and forecast tourism demand. *Decision Support Systems, 123*, 113075. https://doi.org/10.1016/j.dss.2019.113075

Dergiades, T., Mavragani, E., & Pan, B. (2018). Google trends and tourists' arrivals: Emerging biases and proposed corrections. *Tourism Management, 66*, 108-120. https://doi.org/10.1016/j.tourman.2017.10.014

Eichenauer, V. Z., Indergand, R., Martínez, I. Z., & Sax, C. (2022). Obtaining consistent time series from Google trends. *Economic Inquiry, 60*(2), 694-705. https://doi.org/10.1111/ecin.13049

EPRS. (2023). *Rural tourism*. Retrieved from https://www.europarl.europa.eu/RegData/etudes/BRIE/2023/751464/EPRSBRI(2023)751464EN.pdf

Frost, W., Frost, J., Strickland, P., & Maguire, J. S. (2020). Seeking a competitive advantage in wine tourism: Heritage and storytelling at the cellar-door. *International Journal of Hospitality Management, 87*, 102460. https://doi.org/10.1016/j.ijhm.2020.102460

Guaita Martínez, J. M., Martín Martín, J. M., Salinas Fernández, J. A., & Mogorrón-Guerrero, H. (2019). An analysis of the stability of rural tourism as a desired condition for sustainable tourism. *Journal of Business Research, 100*, 165–174. https://doi.org/10.1016/j.jbusres.2019.03.033

Gunter, U., Önder, I., & Gindl, S. (2019). Exploring the predictive ability of LIKES of posts on the Facebook pages of four major city DMOs in Austria. *Tourism Economics, 25*(3), 375-401. https://doi.org/10.1177/1354816618793765

Havranek, T., & Zeynalov, A. (2021). Forecasting tourist arrivals: Google trends meets mixed-frequency data. *Tourism Economics, 27*(1), 129-148. https://doi.org/10.1177/1354816619879584

INE. (2023). *National institute of statistics*. Retrieved from https://www.ine.es

Jepson, D., & Sharpley, R. (2015). More than sense of place? Exploring the emotional dimension of rural tourism expe- riences. *Journal of Sustainable Tourism, 23*(8-9), 1157–1178. https://doi.org/10.1080/09669582.2014.953543

Jun, S.-P., Park, D.-H., & Yeom, J. (2014). The possibility of using search traffic information to explore consumer product attitudes and forecast consumer preference. *Technological Forecasting and Social Change, 86*, 237-253. https://doi.org/10.1016/j.techfore.2013.10.021

Jun, S.-P., Yoo, H. S., & Choi, S. (2018). Ten years of research change using Google trends: From the perspective of big data utilizations and applications. *Technological Forecasting and Social Change, 130*, 69-87. https://doi.org/10.1016/j.techfore.2017.11.009

Kastenholz, E., Marques, C. P., & Carneiro, M. J. (2020). Place attachment through sensory-rich, emotion-generating place experiences in rural tourism. *Journal of Destination Marketing & Management, 17*, 100455. https://doi.org/10.1016/j.jdmm.2020.100455

Komppula, R. (2014). The role of individual entrepreneurs in the development of competitiveness for a rural tourism destination–A case study. *Tourism Management, 40*, 361-371. https://doi.org/10.1016/j.tourman.2013.07.007

Kumar, S., & Valeri, M. (2022). Understanding the relationship among factors influencing rural tourism: A hierarchical approach. *Journal of Organizational Change Management, 35*(2), 385-407. https://doi.org/10.1108/JOCM-01-2021-0006

Lane, B., & Kastenholz, E. (2015). Rural tourism: The evolution of practice and research approaches–towards a new generation concept? *Journal of Sustainable Tourism, 23*(8-9), 1133-1156. https://doi.org/10.1080/09669582.2015.1083997

Li, X., Law, R., Xie, G., & Wang, S. (2021). Review of tourism forecasting research with internet data. *Tourism Management, 83*, 104245. https://doi.org/10.1016/j.tourman.2020.104245

Li, X., Pan, B., Law, R., & Huang, X. (2017). Forecasting tourism demand with composite search index. *Tourism Management, 59*, 57-66. https://doi.org/10.1016/j.tourman.2016.07.005

Liu, C., Dou, X., Li, J., & Cai, L. A. (2020). Analyzing government role in rural tourism development: An empirical investigation from China. *Journal of Rural Studies, 79*, 177-188. https://doi.org/10.1016/j.jrurstud.2020.08.046

Liu, Y.-Y., Tseng, F.-M., & Tseng, Y.-H. (2018). Big data analytics for forecasting tourism destination arrivals with the applied Vector Autoregression model. *Technological Forecasting and Social Change, 130*, 123-134. https://doi.org/10.1016/j.techfore.2018.01.018

Liu, Y., Chiang, J., & Ko, P. (2023). The benefits of tourism for rural community development. *Humanities and Social Sciences Communications, 10*(1), 1-12. https://doi.org/10.1057/s41599-023-01610-4

Martínez, J. M. G., Martín, J. M. M., Fernández, J. A. S., & Mogorrón-Guerrero, H. (2019). An analysis of the stability of rural tourism as a desired condition for sustainable tourism. *Journal of Business Research, 100*, 165-174. https://doi.org/10.1016/j.jbusres.2019.03.033

Nooripoor, M., Khosrowjerdi, M., Rastegari, H., Sharifi, Z., & Bijani, M. (2021). The role of tourism in rural development: Evidence from Iran. *GeoJournal, 86*(4), 1705-1719. https://doi.org/10.1007/s10708-020-10153-z

Önder, I. (2017). Forecasting tourism demand with Google trends: Accuracy comparison of countries versus cities. *International Journal of Tourism Research, 19*(6), 648-660. https://doi.org/10.1002/jtr.2137

Park, S., Lee, J., & Song, W. (2017). Short-term forecasting of Japanese tourist inflow to South Korea using Google trends data. *Journal of Travel & Tourism Marketing, 34*(3), 357-368. https://doi.org/10.1080/10548408.2016.1170651

Rid, W., Ezeuduji, I. O., & Pröbstl-Haider, U. (2014). Segmentation by motivation for rural tourism activities in The Gambia. *Tourism Management, 40*, 102-116. https://doi.org/10.1016/j.tourman.2013.05.006

Rivera, R. (2016). A dynamic linear model to forecast hotel registrations in Puerto Rico using Google trends data. *Tourism Management, 57*, 12-20. https://doi.org/10.1016/j.tourman.2016.04.008

Rocca, L. H. D., & Zielinski, S. (2022). Community-based tourism, social capital, and governance of post-conflict rural tourism destinations: The case of Minca, Sierra Nevada de Santa Marta, Colombia. *Tourism Management Perspectives, 43*, 100985. https://doi.org/10.1016/j.tmp.2022.100985

Rosalina, P. D., Dupre, K., & Wang, Y. (2021). Rural tourism: A systematic literature review on definitions and challenges. *Journal of Hospitality and Tourism Management, 47*, 134-149. https://doi.org/10.1016/j.jhtm.2021.03.001

Rosalina, P. D., Dupre, K., Wang, Y., Putra, I. N. D., & Jin, X. (2023). Rural tourism resource management strategies: A case study of two tourism villages in Bali. *Tourism Management Perspectives, 49*, 101194. https://doi.org/10.1016/j.tmp.2023.101194

Song, H., Witt, S. F., & Li, G. (2008). *The advanced econometrics of tourism demand* (1st ed.). London: Routledge.

UNWTO. (2019). *UNWTO tourism definitions*. Retrieved from https://www.e-unwto.org/doi/epdf/10.18111/9789284420858

UNWTO. (2023). *Economic contribution and SDG*. Retrieved from https://www.unwto.org/

Wang, L.-E., Cheng, S.-K., Zhong, L.-S., Mu, S.-L., Dhruba, B. G., & Ren, G.-Z. (2013). Rural tourism development in China: Principles, models and the future. *Journal of mountain science, 10*, 116-129. https://doi.org/10.1007/s11629-013-2501-3

Whitelaw, P. A., King, B. E., & Tolkach, D. (2014). Protected areas, conservation and tourism–financing the sustainable dream. *Journal of Sustainable Tourism, 22*(4), 584-603. https://doi.org/10.1080/09669582.2013.873445

Wickramasinghe, K., & Ratnasiri, S. (2021). The role of disaggregated search data in improving tourism forecasts: Evidence from Sri Lanka. *Current Issues in Tourism, 24*(19), 2740-2754. https://doi.org/10.1080/13683500.2020.1849049

Wijijayanti, T., Agustina, Y., Winarno, A., Istanti, L. N., & Dharma, B. A. (2020). Rural tourism: A local economic development. *Australasian Accounting, Business and Finance Journal, 14*(1), 5-13. http://dx.doi.org/10.14453/aabfj.v14i1.2

Wilson, S., Fesenmaier, D. R., Fesenmaier, J., & Van Es, J. C. (2001). Factors for success in rural tourism development. *Journal of Travel Research, 40*(2), 132-138. https://doi.org/10.1177/004728750104000203

Xi, C., & Donglai, C. (2022). Application of improved algorithm based on four-dimensional resnet in rural tourism passenger flow prediction. *Journal of Sensors, 2022*, 555–571. https://doi.org/10.1155/2022/9675647

Yang, J., Yang, R., Chen, M.-H., Su, C.-H. J., Zhi, Y., & Xi, J. (2021). Effects of rural revitalization on rural tourism. *Journal of Hospitality and Tourism Management, 47*, 35-45. https://doi.org/10.1016/j.jhtm.2021.02.008

Yin, L. (2020). Forecast without historical data: Objective tourist volume forecast model for newly developed rural tourism areas of China. *Asia Pacific Journal of Tourism Research, 25*(5), 555-571. https://doi.org/10.1080/10941665.2020.1752755