Universitat de les
Illes Balears

Title: Forecasting tourism arrivals with an online search engine data: A study of the Balearic Islands

Master's Thesis

Master's degree in Tourism Management and Planning

(With a speciality/Itinerary in Tourism Management)

at the

UNIVERSITAT DE LES ILLES BALEARS

Abstract

This study explores issues related to the forecasting in revenue management in the Mechanism of predicting tourism arrivals for the Balearic Islands. Specifically, the study uses queries from a web search data (Google Trends) in order to demonstrate the forecasting power of such measures compared to traditional methods. I developed a database formed by the two main tourist volumes, namely, Germany and UK and then, compare each model with its corresponding baseline to figure out whether the Google Trends indicator can increase accuracy of the prediction. After estimating the four different models, I selected the baseline model for Germany, and the alternative for UK. Consequently, Granger causality test indicated a positive causality between variables suggesting good estimating results. Besides, I calculated the Mean Absolute Percentage Errors (MAPE) for each model and the results showed a considerable improvement of the Google Trends models compared to baseline models. The results provide some hints for increasing company efficiency and enhance policy maker decision making.

2.- Introduction

Throughout the last fifty years the boom of the tourism activity affected the Balearic Islands significantly. Nowadays, this sector far from being considered as declining is on the stagnation phase according to the expanded version of the tourism lifecycle of Butler (1980).

In fact, within the Balearics the tourism industry is the main source of the regional GDP. The Balearics GDP has grown 3.2% during the last year (http://www.ibestat.cat/ibestat/inici) and it is mainly boosted by the influence of the tourism activity. This enormous growth implied a widespread increased of the number of tourism arrivals, and hence it demonstrates the urgency for proper prediction, especially near-term prediction, in order to correctly allocate resources and managing tourist flows.

Thus, most of the current literature of the Balearic Islands has focused on analyzing the environmental problems directly derived from the tourism activity and how to solve them by implementing “Eco taxes” or other instruments (e.g. Aguiló, Riera and Rosselló, 2005; Palmer and Riera, 2003). Meanwhile others have studied the environmental innovation as a source of increase competitiveness (e.g. Jacob, Florido and Aguiló, 2010). However, Alvarez-Diaz, Mateu-Sbert and Rosselló-Nadal (2009) sought how to forecast the monthly tourism demand on UK and Germany arrivals for the Balearic Islands but none of them did it by using search engine data. In addition, if companies and governments are better able to forecast the number of tourists, they will be willing to assess more efficiently resources.

Actually, the tourist volume forecasting is based on a bundle of techniques such as statistical or econometric models that rely on historical data to forecast future tourist activities by assuming ceteris-paribus the economic environment. Therefore, these methods might not be so accurate since they primarily focus on long-term horizons such as yearly or quarterly, instead of monthly either weekly data (Yang *et al.,* 2015).

The globalization jointly with the fast evolution of the ICT technologies has led hundreds of millions of different search queries by tourists. In fact, these queries reflect the possible customers’ trends, but also offering the possibility to forecast their future behavior. This study will use a Web search query to generate data useful for forecasting the numbers of visitors coming to the Balearic Islands. Specifically, Google will be the search engine to be used in the project. Particularly, query data on visitors will be generated by using Google Trends.

The revenue management appeared during the 70’s in the U.S. travel industry. However, its importance and relevance has dramatically increased due to the higher competition among companies. Since, companies compete more; they need to be more efficient in allocating their resources (i.e. physical, human and financial resources) and thus increase their market share in order to survive.

Nevertheless, our analysis will be focused on the revenue management area. One of its main components is the forecasting because during the whole process the demand, the supply and the overbooking controls must be predicted at different levels. For example, forecasting the tourism demand would imply to figure out the number of rooms that are going to be booked for the next week. Moreover, forecasting the supply, namely late or early checkouts, is useful to assess the number of rooms that can be sold. And lastly, to correctly predict the number of cancellations and no-shows is vital in order to set an adequate overbooking policy. Again, Vinod (2004; pp. 183) said "*high forecast errors will result in conservative inventory controls and increase the likelihood of revenue dilution*".

2.1. Predicting tourism demand

Traditionally, since its origin the usual framework for predicting tourism demand was to implement time series model and its variations. Actually, these models are well-established and, in fact, are somehow better than other methods (Song, Witt and Li, 2008).The development of ICT technology has increased the use of non-conventional methods for predicting tourism demand, such as artificial intelligence methods. In fact, one interesting case is the study performed by Alvarez-Diaz *et al.* (2009). They used Genetic Program (GP) to predict the monthly arrivals of UK and Germany to the Balearic Islands. Thus, the study compared the performance of the GP model against different univariate models such as no-change model, Moving Average and ARIMA. Finally, they concluded that GP can be used as a source for forecasting tourism arrivals since the GP gets better estimations for the case of the German demand.

2.2. Predicting tourism demand with search engine data

Search engine data has been recently used as a source for gathering information and forecasting socioeconomic activities in different knowledge branches such as medicine (Althouse, Ng and Cummings, 2011; Ginsberg, Mohebbi, Patel, Brammer, Smolinski and Brillian, 2009).

In fact, the most common search engine used is Google Trends due to its potential to inform strategic decision making in tourism destinations. A lot of research is recently been carried out, for instance, Jackman and Naitram (2015) analyzed whether the inclusion of Google Trends index within a support vector regressions (SVR) could enhance accuracy of forecasting for the Barbados Island tourist arrivals. Indeed, they refer to forecasting as nowcasting based on Castle, Fawnett and Hendry (2009), "*nowcasting, in its most basic form, can be summarized as predicting the present and sometimes the recent past*" as a way for predicting the incoming future. Lastly, they both concluded that including the Google Trends index increase the accuracy of estimations by means of a lower MAPE for the predictions of UK and Canadian tourists, whilst not for the US tourist arrivals. Rivera (2016) considered the information gathered by Google Trends as a source for predicting the number of hotel nonresident registrations in Puerto Rico. He showed a positive result in conducting the analysis; however, some problems arose from the model and finally concluded that for the short term the Holt-Winter forecast resulted in smaller forecast errors, while for the long term periods their model using Google Trends index outperformed other models. Bangwayo-Skeete and Skeete (2015) sought whether the introduction of a new indicator based on Google Trends could improve the accurateness efficiency of current forecasting models for Caribbean destinations using data on US, UK and Canadian travellers. Finally, both concluded that AR-MIDAS model, which includes the queries in Google, gave superior predictions to SARIMA and AR models in terms of accuracy (e.g. RMSE and MAPE criteria).

Overall, the accuracy of the forecast becomes a key issue in determining company's and DMO strategy in order to promote a sustainable tourism development. As mentioned before, enhancing the estimators' prediction power would yield to a higher efficiency, and thus greater welfare. Therefore, testing the accuracy of current models versus new techniques such Google Trends source could yield to a better comprehension of the revenue management and, management in general. In addition, the limitation of the traditional predictive methods lies on the time lag between the collection and publication of the data and also, databases can be based on insufficient samples, hence providing useless predictions (Huang *et al.*, 2016). I therefore propose to compare the fitness and prediction of the current forecasting models versus Google search data in forecasting the tourist volume for the Balearic Island. Additionally, based on Yang *et al.* (2015), I would implement a systematic way to better select queries for predicting.

3.- Purpose of the project

Therefore, in this study we proposed a baseline project based on the search engine, named Google Trends, which will be compared with the actual forecasting models with their equivalent time counterpart in order to assess the validity of the tourism arrivals forecasting to the Balearic Island in shorter periods.

The implications of our study contribute to the existent literature of the revenue management in two ways. First, confirming the statistical significance and accurateness of search engine tools such as Google Trends, in predicting the tourism arrivals to a well-known tourism destination like the Balearic Islands. Second, the managerial decision that companies either destination management organizations (DMO) could perform whether they better forecast the tourism demand, in terms of higher efficiency and effectiveness of resources

4.- Methodology & work plan

4.1. Empirical testing

The Balearic Islands was chosen as destination for testing our empirical model. It is well known that since the middle 60s the attractive of the Balearic Islands had increased dramatically. Especially during these last years the number of tourist arrivals has boosted, in 2014, the Balearic Island received over 13.5 million overnight tourists, with 11,348,000 tourists (almost 84%) of them being foreigner visitors. In fact, most of the tourist arrivals are concentrated in two major countries Germany and United Kingdom with 4.1 and 3.4 millions of tourist arrivals respectively. Overall, these countries represent 7.52 million of arrivals implying more than 55% of the whole tourist demand for the year 2014 (Ibestat, 2016).

4.2. Data source

Google Trends is a search engine tool that provides an index of the different Google queries sorted by geographic location and category. Thus, the database will be directly extracted from Google Trends since it represents the search engine with the largest market share in Germany and in the U.K., particularly 97.09% and roughly 90% respectively (Kennedy and Hauksson, 2012).

Moreover, Google Trends reports a query index which displays frequently how often a particular query has been searched compared to the total search volume from different countries and languages; actually it does not report the raw level of total queries. In Choi and Varian (2009) explain the entire process of how Google Trends creates the index, and how it began at January 2004. Since that year, the numbers indicate the percentage of variance from the query share in 2004

4.3. Selection of search queries

In order to facilitate the analysis, I concentrated the search queries database in the main tourism destination of the Balearic Islands, which is Mallorca. In 2014 Mallorca gathered 9.6 million of tourist arrivals, who came principally from Germany and U.K., 3.7 and roughly 2.2 million respectively (Ibestat, 2016). So, as it is stated above, from Yang *et al.* (2015) results the four stage process used in my study to select the different search queries and to create the index that will be used in the different models.

(1) Initially 20 basic search queries were selected based on the different trip aspect posed by Pan *et al.* (2006) including 10 search queries in German and another 10 search queries in English.

(2) Afterwards, the 20 queries were introduced in Google Trends as seed queries and retrieved the related queries. Then, I chose the related queries and repeat the procedure for second and third rounds. The total number of queries converged to 134, however only 100 queries remained in the database after duplications were removed. In fact, the UK database is composed by 49 search queries and the German database is formed by 51 search queries.

(3) I have calculated the Pearson correlation coefficient between Mallorca monthly visitor volume and each of the search queries with different lag periods. In fact, for each query I calculated eight correlation coefficients of 0-7 months ahead respectively. Moreover, from the two datasets (i.e. Germany and UK) I selected 17 queries from UK, and 17 queries from Germany. In order to calculate an appropriate number of search queries, I used 0.65 as the threshold for the correlation between visitors' volume and Google Trends data (Yang et al., 2015). The reason of that threshold is the following. First, if we set a threshold below 0.65, let us imagine that I choose a threshold of 0.60, then 43 keywords for the UK and 38 keywords for Germany would be selected and it might reduce the parsimony and generalizability of the models. Second, if I increase the threshold to 0.70 the number of keywords would be 10 for the UK and 9 for Germany and therefore, it will result in a low forecasting accuracy. Hence, the selection of the threshold represents the trade-off between forecasting accuracy and model parsimony.

(4) Since the purpose is to forecast the future tourist volume, based on Yang *et al.* (2015) I only select queries with, at least, one lag previous to the arrival. Therefore, 8 queries with 1 lag were chosen for Germany and 11 queries with 1 and 7 lags were chosen for UK as Google Trends predictors.

The data reveals interesting information about the travel behaviour of both countries. First, for the German tourists the lags of the maximum correlation coefficient varied mainly from 0 to 1, distributed similarly across the sample. Indeed, they are concentrated in the transportation and destination such as Mallorca airport arrival, Mallorca departures, Alcudia Mallorca, Mallorca beach. Nevertheless, it is important to point out that the weather queries were also relevant for the German tourists. Second, the highest correlated lags for the UK travelers. It varies from 7 to 0 lags, being 1 the most common. Furthermore, it gathers information about destination (e.g. Calas de Mallorca, Palma de Mallorca, and Palma Mallorca), transportation (e.g. Mallorca travel, Mallorca airport, Palma departures) and weather information (e.g. Weather in Majorca, Weather in Alcudia).

4.4. Search data index

In order to develop the different models I aggregated search data by using PCA analysis. Firstly, I computed Cronbach's Alpha to validate and testing the internal consistency of the PCA analysis. In fact, I obtained a value of 0.805 and 0.951 for the UK and German data respectively (recommended value of 0.7). Then, a PCA analysis was calculated for each database.

In the case of Germany the main indexes confirm the appropriateness of using this analysis (Delgado-Verde *et al*., 2011), since the KMO index had a value of 0.88 (higher than 0.6 that is the value recommended); the Bartlett test was significant at a level lower than 0.05 (0.000); the extraction column of the commonality showed high values between 0.70 and 0.91 that can be interpreted as the factor analysis adjust level. In addition, I applied Varimax orthogonal rotation, all items that had a load higher than 0.85 (items with a load lower than 0.4 were excluded from the table), finding one factor that deeply capture the extent and the degree of the relationship. Finally, the percentage of accumulated explained variance for the factor was 83.26 percent, being higher than the proposed value for social science: 60 percent (Hair, Anderson, Tatham, and Black, 2004).

I further calculated the PCA analysis for the UK data, and the results confirmed the appropriateness of implementing this framework. The KMO test showed a value of 0.893; the Barlett test was significant at a level lower than 0.05 (0.000); besides the communalities' extraction column showed high values between 0.705 and 0.942 indicating the factor adjusting level. Furthermore, Varimax orthogonal rotation was implemented obtaining item loads higher than 0.70, finding two factors that can be used to capture most of the information of the variables. Indeed, the accumulated explained variance by the two factors was 81.35 percent.

5.- Analysis and results

Considering the PCA conclusions I constructed two different time series models based on Google Trends data for Germany and UK. The dependent variable in the three models is T_{t1} which denotes the Mallorca monthly tourist volume, from August 2004 to June 2016.

$$\text{Log}\, T_{t1} = c_0 + \beta_1 \text{Log}\, T_{t1}\,(-12) + u_t \qquad (1)$$

$$\text{Log}\, T_{t1} = c_0 + \beta_1 \text{Log}\, Ger_{t1} + u_t \qquad (2)$$

$$\text{Log}\, T_{t1} = c_0 + \beta_1 \text{Log}\, UK_{t1}^1 + \beta_2 \text{Log}\, UK_{t1}^2 + u_t \qquad (3)$$

Eq. (1) represents the baseline model which uses historical tourist volume data to predict the actual tourism arrivals. Since, the tourism demand presents seasonality features I decided to use as predictor the 12 periods T_{t1}. Moreover, Eq. (2) showed the first model with the PCA component derived from the German database. In addition, Eq. (3) represents the second model with PCA components using UK data from Google Trends, being UK_{t1}^1 the first weight index and UK_{t1}^2 the second weight index. Furthermore, the forecasting models of each country have been compared with its corresponding baseline model.With equations (1)-(3) I analyzed the correlogram and the partial correlogram of each independent variable in order to figure out whether the variable follows a stationary process or not. Moreover, augmented Dickey-Fuller (ADF) unit-roots test were applied to all independent variables in each model (see Table 6). Only the two dependent variables of the baseline models (i.e. Log$T_{t1}^{GER}$for Germany and LogT_{t1}^{UK} for UK) needed to be modified in order to transform to a stationary process. Moreover, I computed the ADF tests for the residuals of each regression in order to check for any co-integration relationship. The evidence shows that there exists a positive co-integration relationship between the exogenous and endogenous variable of each model. In fact, the co-integration relationship implies that there exists a long-term relation between variables, so that when both variables grow in time T, they both do it in a totally synchronized form such that the error term between the variables does not increase.

On the one hand this might support the Granger causality analysis and, on the other hand my modelling process based on ARMAX models (Autoregressive Moving Average with External Variables). In all the four models, I employed data from August 2004, to December 2015, I omitted the last six periods in order to out-sample forecast. I created a baseline model for each database, namely I developed a baseline model for Germany L_{b1} and a baseline model for UK L_{b2}. Then, I estimated an ARMAX model for Germany and UK and I compared them with their corresponding baseline model.

5.1. ARMAX results for Germany

In L_{b1} all variables were significant at a 0.01 level, although the exogenous variable, in this case Log$T_{t1}(-12)$, presented a coefficient close to zero (-0.0001). In addition, the constant, moving averages of order 1 and 12, and autoregressive of order 1 and 2 were statistically significant at a 0.01 level. The expression of the model is represented at equation 4

$$\begin{cases} \text{Log}\, T_{t1} = 12.375 - 0.0001\, \text{Log}\, T_{t1}\,(-12) + u_t \\ u_t = 1.429\, u_{t-1} - 0.752\, u_{t-2} + \varepsilon_t - 0.466\, \varepsilon_{t-1} + 0.701\, \varepsilon_{t-12} \end{cases} \qquad (4)$$

The ARMAX model for Germany that best fitted the data was an ARMAX (1, 12) meaning that the error term presented an autoregressive part of order 1 and a moving average of order 12. The independent variable and all the autoregressive and moving average coefficients were statistically significant at a 0.01 level. The positive coefficient of the PCA index suggested a correlation between web search data and Mallorca tourist volumes.

5.2. ARMAX results for the UK

The ARMAX model with the UK principal component factors is defined in equation 7. In the model, all coefficients were statistically significant at a 1% level, although the first component was not (i.e. UK_{t1}^{1}) neither at a 10% of significance. The second component was statistically significant and presented a positive coefficient indicating a positive correlation between the weather search information and the Mallorca visitor arrivals.

$$\begin{cases} \operatorname{Log} T_{t1} = 11.812 - 0.069\, UK_{t1}^{1} + 0.064\, UK_{t1}^{2} + u_t \\ u_t = 0.989\, u_{t-12} + \varepsilon_t + 0.414\, \varepsilon_{t-1} + 0.180\, \varepsilon_{t-2} - 0.279\, \varepsilon_{t-12} \end{cases} \quad (7)$$

I tested for unit-root implementing an ADF in the residuals and both models lead to a stationary conclusion at a 0.01 level. This confirmed the co-integration association between dependent and independent variables, and the idea that there is a long-term association between variables. Based on the various tests like Log likelihood, AIC and BIC criteria I selected the UK model since it presents better estimation results.

5.3. Granger causality analysis

The existence of a correlation, either positive or negative, among two variables does directly imply causality between them. Indeed, it does mean that one variable causes the fluctuations of the other. Theses causes and consequences might come from a spurious origin. Thus, the Granger causality tests allow us to analyze whether a variable X causes variable Y. In fact, under null hypothesis there is no causality relationship between variables and under alternative hypothesis otherwise.

Consequently, I tested Granger causality for our two ARMAX models with principal component indexes (i.e. G1 and UK1). Yet, in the case of Germany it was preferable to select the baseline model I wanted to analyze the relationship between variables to figure out the possible existence of causality. Due to the great sensitivity to the lag order, I previously considered five test criteria for the selection of the lag order: LR (Likelihood Ratio Test), FPE (Final Prediction Error Criterion Minimum), AIC (Akaike Information Criterion), SBIC (Schwarz Information Criterion) and HQ (Hannan-Quinn Information Criteria).

The different criteria lead to the selection of the same lag order of 12. This order was used in the Granger causality test. Furthermore, the results of the analysis showed that for the G1 model Ger_{t1} and $\operatorname{Log} T_{t1}$ Granger caused each other, that means, Google Trends data can predict Mallorca tourism volume and vice versa. However, for the UK1 model UK_{t1}^{2} and $\operatorname{Log} T_{t1}$ I only found out a positive Granger causality for the UK Google Trends data implying that Google Trends can be used as a predictor of UK tourism volumes, although I did not find support for the opposite causal relationship, since the p.value is larger than 0.05 (0.7303).

5.4. Forecasting with web search data

In order to test the predictive power and accuracy of the different models, I dropped from the training set the last six months from January 2016 until June 2016 for testing and compare the results. I compared the actual value of the dependent variable with its corresponding predicted value, and then calculated the percentage of error and lastly, the MAPE (Mean Absolute Percentage Error).

The results indicated that both models, G1 and UK1, predicted 6 months of the Mallorca visitors' volume more accurately than their corresponding baseline models. In fact, the G1 model improves the results in a roughly 2%, although in the UK case the UK1 model enhances the prediction by reducing the error in a 5%.

In addition, the actual value of the dependent variable (i.e. the natural log of the tourism arrivals) and also, the fitted value of the two estimated models. Hence, the results pointed out that model UK1 outperforms model G1 in fitting the data, since it seemed to present less variance.

6.- Conclusions

My article makes two principal contributions in the existing forecasting literature. First, this is the first study that forecast tourism arrivals in the Balearic Islands by implement search engine methods such as Google Trends, and the results supported the idea that the Google Trends index enhances forecasting accuracy. In fact, this can pose the base for further research on that topic since it is interesting to reveal the structure of the "nowcasting" models. Second, I supported the methodology for query selection to better fit and predict visitors' volume suggested by Yang *et al.* (2015), however, I applied slightly variations to the initial methodology.

6.1. Implications

Our results reveal relevant implications for managerial decisions and for policy maker decisions. This new forecasting approach can influence managerial decisions mainly in the tourism and hospitality services by developing a new framework for monitoring and tracking short-term information about the consumers' demand. For instance in the case of a hotel, the capacity is limited for the short-term, since the hotel cannot enlarge the capacity from one day to another. Thus a room which is not occupied implies directly a loss to the hotelier, and therefore a decrease in the profits generated. Whether the hotel company is able to better adapt the demand, by increasing the forecasting accuracy including web search data into their forecasting models as I have shown in this paper, then the hotel will be able to improve its efficiency.

Indeed, our results pose some hints on how to solve some problems that revenue management tries to answer. For example, the forecasting part of the revenue management is crucial for determining overbooking strategies, pricing strategies, and so on. Hence, compared to traditional models of monitoring visitor numbers, the predictive power of our models based on web search data is much higher.

Moreover, policy makers could use web search queries of a particular region to release a forecasted tourists' index and companies can use it as a benchmark for local tourism and hospitality companies to measure their performance. For instance, whether the tourists' index showed a 30% of increase for a certain area in a month, yet if the reservation volume of that area only increases a 15%it should not be considered as a good result since there is still a 15% gap of reservation volume that will be lost. Furthermore, enhancing the accuracy of the estimating models will lead to a better prediction of the tourism demand and hence an increase in the efficiency of the resources managed by policy makers. For instance, in the Balearic Islands the overcrowded seasonal months (i.e. June, July and August) imply the need for better handle the tourism activity, such as the environmental management. Lastly, policy makers can use the forecasted demands to predict the possible income generated by imposing a tourism tax such as the Ecotax.

6.2. Limitations and further research

Nevertheless our study presents some limitations that can pose a milestone for future research in this topic. The first one is that I only focused in Balearic Islands tourism arrivals; therefore the ability to generalize the conclusions is limited. Another important limitation that I ad hoc experimentation reveals that there exists a gap between Google Trends data and travel agency queries, at least, for the two countries that I selected. In fact, this might pose a threat to our results since they do not gather information regarding travel agency companies that play a key role in the travel and plan process of these two tourism countries. Further research in this last issue must be studied in order to clarify the significance of the relationship, and how it can influence the estimation results.

7.- Bibliography

Aguiló, E., Riera, A., & Jaume, R. (2005). The short-term price effect of a tourist tax through a dynamic demand model: The case of the Balearic Islands, *Tourism Management,* Vol. 26 (3), pp. 359-365.

Althouse, B.M., Ng, Y. Y., & Cummings, D.A. (2011). Prediction of dengue incidence using search query surveillance, *PLoS Neglected Tropical Diseases,* Vol. 5 (8), e1258.

Alvarez-Diaz, M., Mateu-Sbert, J.,& Rossello-Nadal, J. (2009). Forecasting tourist arrivals to Balearic Island using genetic programming, *International Journal of Computational Economics and Econometrics,* Vol. 1(1), pp. 65-75.

Álvarez-Díaz, M.,& Rosselló-Nadal, J. (2010). Forecasting British tourist arrivals in the Balearic Islands using meteorological variables, *Tourism Economics,* Vol. 16 (1), pp. 153-168.

Arenoe, B., Van der Rest, J.P., & Kattuman, P. (2015). Game theoretic pricing models in hotel revenue management: An equilibrium choice-based conjoint analysis approach, Vol. 51, pp. 96-102.

Bangwayo-Skeete, P.F., & Skeete, R.W. (2015). Can Google data improve the forecasting performance of tourist arrivals? Mixed-data sampling approach, *Tourism Management,* Vol. 46, pp. 454-464.

Butler, R. (1980). The concept of a Tourist Area Cycle of Evolution: Implications for Management of Resources, *Canadian Geographer,* Vol. 24, pp. 5-12.

Canestrelli, E., & Costa, P. (1991). Tourist capacity: A fuzzy approach, *Annals of Tourism Research,* Vol. 18 (2), pp. 295-311.

Castle, J.L., Fawcett, N.W., & Hendry, D.F. (2009). Nowcasting is not just contemporaneous forecasting, *National Institute Economic Review*, Vol. 210 (1), pp. 71–89.

Choi, H., & Varian, H. (2009). Predicting the Present with Google Trends, (Google), Technical report.

Chu, F.L. (2009). Forecasting tourism demand with ARMA-based methods, *Tourism Management,* Vol. 30 (5), pp. 740-751.

Delgado-Verde, M., Martín-de-Castro, G., Navas López, J.E., & Cruz-González, J. (2011a). Capital social, capital relacional e innovación tecnológica. Una aplicación al sector manufacturero español de alta y media-alta tecnología, *Cuadernos de Economía y Dirección de la Empresa*, 14, pp. 207-221.

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,* Vol. 457 (7232), pp. 1012-1014.

Gounopoulos, D., Petmezas, D., & Santamaria, D. (2012). Forecasting tourist arrivals in Greece and the impact of macroeconomic shocks from the countries of tourists' origin, *Annals of Tourism Research,* Vol. 39 (2), pp. 641-666.

Hair, J.F. Jr, Anderson, R.E., Tatham, R.L. & Black, W.C. (2004). Análisis multivariante, 5th ed.,Pearson-Prentice Hall, Madrid.

Huang, X., Zhang, L., & Ding, Y. (2016). The Baidu Index: Uses in predicting tourism flows: A case study of the Forbidden City, *Tourism Management*, In press.

Ibestat, Institut d'Estadística de les Illes Balears (2016). *Estadísticas Economía, turismo, flujo de turistas.* Retrieved from http://www.ibestat.cat/ibestat/estadistiques/economia/turisme/fluxe-turistes-frontur/043d7774-cd6c-4363-929a-703aaa0cb9e0. Accessed 26.07.16

Jackman, M., & Naitram, S. (2015). Nowcasting tourist arrivals in Barbados- just Google it!, *Tourism Economics,* Vol. 21 (6), pp. 1309-1313.

Jacob, M., Florido, C. & Aguiló, E. (2010). Environmental Innovation as a competitiveness factor in the Balearic Islands,*Tourism Economics,* Vol.16 (3), pp. 755-764

Kennedy, A.F., & Hauksson, K.M. (2012). *Global Search Engine Marketing: Fine-tuning Your International Search Engine Results*. USA, Indiana: Pearson.

Li, G., Song, H., & Witt, S. F. (2005). Recent developments in econometric modelling and forecasting, *Journal of Travel Research*, Vol. 44, pp. 82–99.

Palmer, T.,& Riera, A. (2003). Tourism and environmental taxes. With special reference to the "Balearic Ecotax", *Tourism Management,* Vol. 24(6), pp. 665-674.

Pan, B., Litvin, S.W., & Goldman, H. (2006). Real Users, Real Trips, and Real Queries: An analysis of Destination Search on a Search Engine, Annual Conference of Travel and Tourism Research Association, Ireland.

Schwartz, Z., Stewart, W., & Backlund, E. (2012). Visitation at capacity-constrained tourism destinations: Exploring revenue management at a national park, *Tourism Management,* Vol. 33 (3), pp. 500-508.

Shen, S., Li, G., & Song, H. (2011). Combination forecast of international tourism demand, *Annals of Tourism Research,* Vol. 38 (1), pp. 72-89.

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

Song, H., Witt, S.F., & Li, G. (2008). *The advanced econometrics of tourism demand.* New York: Routledge

Vinod, B. (2004). Unlocking the value of revenue management in the hotel industry, *Journal of Revenue and Pricing Management,* Vol. 3 (2), pp. 178-190.

Yang, X., Pan, B., Evans, J., & Lv, B. (2015). Forecasting Chinese tourist volume with search engine data, *Tourism Management,* Vol. 46, pp. 386-397.