“Can Internet Searches Forecast Tourism Inflows?”

Concha Artola (Banco de España), Fernando Pinto (Universidad de Salamanca) and Pablo de Pedraza (University of Amsterdam)

1 Introduction

The popularity of the Internet has given rise to a whole range of new activities and has changed the way we go about traditional activities. We increasingly read the press online, look for medical information on the Web before –and after– visiting the doctor, buy products and services online (books, music, air tickets, etc.), participate in social networks, conduct operations with banks and government, write e-mails, make telephone calls and watch TV online. This myriad of activities leaves tracks on the Web, generating an enormous volume of information on products, people, institutions, purchase intentions, voting intentions, economic activities and so on. This is what has come to be known as **big data,** a body of information so extensive and varied that it cannot be processed by software used in traditional databases. In the words of Gary King, the director of Harvard's Institute for Qualitative Social Science: "It´s a revolution, we are really just getting under way. But the march of quantification, made possible by enormous new sources of data, will sweep through academia, business and government. There is no area that is going to be untouched" [1].

The wealth of information in real time nicely suits the needs from central banks and other policy makers –which constantly require timely economic indicators in order to assess the state of the economy and implement policies accordingly. In order to do so they typically rely on official statistics (e.g. on employment, retail sales, house sales), as well as soft data such as business or consumer confidence surveys. The problem with the former is that their compilation requires time: up to several weeks in many instances, which is why Central Banks often gather information on the state of the economy through other more informal sources, such as market intelligence obtained through contacts with industry representatives (e.g. This is the case of Business Outlook Survey of the Bank of Canada , the Federal Reserve Beige Book, the "economic intelligence" gathered by the Reserve Bank of Australia or the Corporate telephone Survey run by the European Central Bank .

In this paper we focus on the tourist industry given that Spain is one of the world's main tourist destinations, and the developed country with the greatest dependence on tourism in the world. We construct an indicator based on Internet searches made by individuals. The idea is that when planning a holiday, people often use the

Internet [2] in order to look for information about their destination (flights, hotels, amenities, weather, etc.). These searches can provide a valuable leading indicator of actual trips.

The article is organised as follows. Section 2 reviews briefly the main trends of the tourism industry in Spain. Section 3 describes the data currently used to forecast tourist visits to Spain, and how to approach the online searching behaviour of potential travellers including a short review of the literature. Section 4 describes and compares our models -which include an indicator measuring online searches- with conventional short-term models. Section 5 discusses some of the limitations of this indicator as well as certain sources of errors and shortcomings. While the improvement of forecasting is really impressive until 2012, the performance of our models worsens in the following years. Section 6 concludes and illustrates why further research is warranted to better understand how Internet search activities translate (or not) into real economic variables.

2 Main trends of the tourism industry in Spain

Spain is one of the world's main tourist destinations (third in terms of arrivals and second in terms of tourism receipts). The importance of the tourist industry for the Spanish economy has been well known for decades, in fact, Spain is the developed country with the greatest dependence on tourism in the world: more than 10% of its GDP and almost 12% of employment are linked to this industry. (Table 1).

Tourism in Spain in Perspective **TABLE 1**

	International arrivals 2013 (million)	Tourism receipts 2013 ($bn)	Tourism as % of GDP 2013	Tourism as % of employment 2013
France (*)	83,0	56,1	7,0	7,1
United States	69,8	139,6	2,8	na
Spain	60,7	60,4	10,8	11,8
China	55,7	51,7	4,0	na
Italy	47,7	43,9	3,7	5,6
Turkey	37,8	28,0	2,8	na
Germany	31,5	41,2	4,4	na
United Kingdom	31,2	40,6	na	5,4
Russia	25,7	12,0	2,3	na
Thailand	22,4	42,1	na	na

SOURCES: World Tourism Organisation (2013) for the first two colums and OECD

Tourism Trends and Policies (2014) for the the last two colums.
(*) Data are for year 2012

The main customers of the Spanish tourist industry are the United Kingdom, Germany and France. After peaking in 2007, tourist inflows declined in the next couple of years; in 2011 the level of entries was 4% below its peak. Only in 2013 have the inflows surpassed the previous maximum (Table 2).

Tourism inflows by country of residence **TABLE 2**

	2007 (million persons)	2013 (million persons)	2013 breakdown by country of residence	2013/2007 accumulated growth
Germany	10,1	9,85	16,3%	-2,5%
France	9,0	9,5	15,7%	5,6%
United Kingdom	16,3	14,32	23,6%	-12,1%
Scandinavian countries (a)	3,4	4,87	8,0%	43,2%
Italy	3,6	3,2	5,3%	-11,1%
Ireland	1,63	1,28	2,1%	-21,5%
Portugal	2,41	1,68	2,8%	-30,3%
Benelux	4,2	4,5	7,4%	7,1%
Rest of Europe (b)	4,2	6,0	9,9%	42,9%
Rest of the World	3,7	5,38	8,9%	45,4%
Total	58,54	60,58	100,0%	3,5%

SOURCE: Instituto de Estudios Turisticos (IET) a. Scandinavian countries: Denmark, Finland, Norway and Sweden. b. Including Russia.

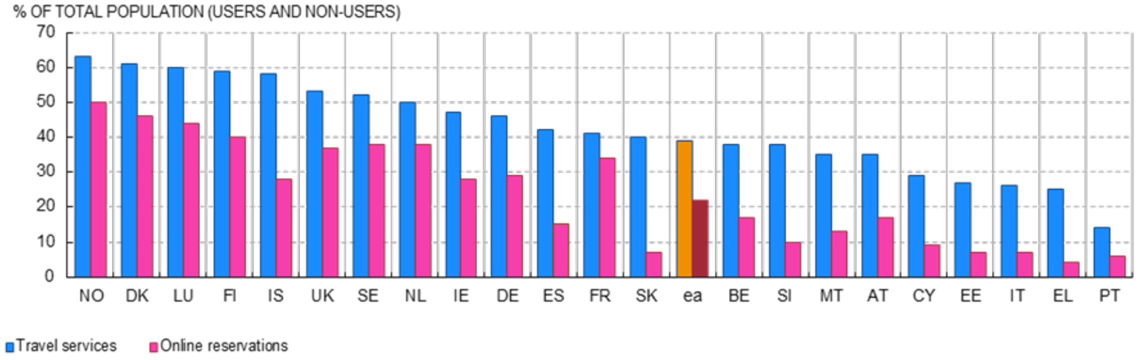
Over the period 2007- 2013 there was a decline in the number of tourists from Germany (-2.5%) and a deeper fall in the number of British tourists (-12%). The initial contraction in the number of French tourists in the first years of the crisis has reverted, and in 2013 its number surpassed the 2007 peak. The number of tourists from the European countries most affected by the crisis –Italy, Ireland and Portugal- has declined the most, while the inflows from the Scandinavian countries (Denmark, Finland, Norway and Sweden) have increased sharply. Tourism from the rest of Europe and other non-European countries brought 11.4 million people to Spain in 2013, partly offsetting the fall-off in tourist flows from its traditional customers.

3 The online searching behaviour of potential travellers: construction of indicators of online searches

Seminal papers using Google trends to forecast economic variables focused on forecasting unemployment claims (Choi and Varian 2009a, Askitas and Zimmerman 2009). Following Choi and Varian (2009b), who showed how to use searches to forecast travel to Hong Kong, other studies have focused on other tourists destinations, Saidi, Scacciavillani and Ali 2010, on Dubai and Jackman and Naitram (2013) on Barbados among others). Several studies have been published by researchers working for Central Banks whose main objective is to provide complementary tools to policy makers: Suhoy (2009) tests predictive ability of queries for real growth of industrial production, retail trade, trade and services revenue, consumer imports and service exports and unemployment rates in Israel, Carrière-Swallow and Labbé (2013) use google trends to improve now-casting models for automobile sales in Chile, McLaren and Shanbhogue (2011) examine the use of online searches for labour and housing markets in the United Kingdom, Artola and Galan (2012) focus on British tourist inflows to Spain and, Matsumoto, Matsumura and Shiraki (2013) analyze travel consumption in Japan.

Evidence supporting the use of the internet queries data to study tourism inflows is provided by two surveys published by Eurostat and the European Commission. According to the *EU Survey on Information and Communication Technologies (ICT) usage in households and by individuals*, 39% of residents in the euro area engage in online activities related to travel, while 22% make online purchases of this type of service. The gap therefore stands at 17 pp. [3] and the size of this discrepancy varies greatly from one country to another (Chart 1), although the two statistics offer a similar message: the correlation between them is 0.903. The reasons behind discrepancies are diverse: internet penetration, consumers' scant trust in online security when making purchases, lower development of online sales by companies operating in the European market. In all the three countries originating the largest flows of tourist unto Spain (United Kingdom, France and Germany), the use of the internet related to travel activities is above the euro area average.

TRAVEL SERVICES AND ONLINE RESERVATIONS (A) **CHART 1**



SOURCE: Eur(a) Travel services include the use of the Internet to gather information and to buy goods and services relating to holiday travel/accommodation. Online reservations comprise those made in the past 12 monthsostat (2010 Survey).

The "Survey on the attitudes of Europeans towards tourism", published by the European Commission provides detailed information on Europeans' travel uses and practices, the percentage of the population who browse Internet as the main source of tourist information increased from 38% in 2009 to 45% in 2011 Browsing the Web stands second only behind recommendations from friends and colleagues (the favourite source for 58% of Europeans) and well ahead of personal experience (21%) and traditional travel agencies (21%) In short, almost half of European citizens seek information on travel and holidays primarily on the Internet. Google is the most popular search engine in Europe, and starting in 2004 has developed a tool, called Google trends, which publicly provides data on the search intensity of different keywords.

In this paper we propose to exploit the fact that individuals make an intensive use of searches when planning their vacation, and use the information provided by Google trends on these searches in order to predict the actual number of visitors for holiday purposes. Two caveats should be mentioned beforehand. First, the set of individuals engaging in online searches for travel purposes are just a fraction of the population of travellers, in which younger and more educated segments of the population are likely to be overrepresented. Second, the “search intensity” indicator provided by Google trends only shows the relative search volume of key words or phrases with respect to the total searches done in a certain geographical area and time period. A decline in the index value for a particular keyword does not necessarily mean that the absolute volume of searches on that particular keyword has declined; it is sufficient if it has increased by less than the total volume of searches from that particular location and time period [4].

We obtain a query index on the relative interest in travelling to Spain for holidays for each of the three aforementioned countries providing the largest flows of tourists into Spain. The key words we use are “Spain Holiday” focusing on searches originating in the United Kingdom, “Vacances Espagne” for searches originating in France and “Spanien Urlaub” for those searches originating in Germany [5] . To further check the robustness of this approach we obtain a query index on the term “Spain” (for searches originating in the UK), “Spanien” for searches originating in Germany and Espagne for searches originating in France) under the category “Travel”. The Google index is provided weekly and is averaged to obtain a monthly series. When a week overlaps across two months it is assigned to the month with the highest number of days laid. For each of the three countries, we compare [Chart 2] actual tourist inflows with the query index provided by Google trends. The time series for tourist inflows is published monthly by the Institute of Tourism Studies (“Instituto de Estudios Turisticos”) which compiles a monthly survey (“Movimientos turisticos en fronteras”) by asking a sample of travellers about their country of residence and the purpose of the trip, thus making it possible to identify the number of tourists per month and country of origin

Chart 2 compares tourist inflows from each country with the corresponding Google query index. The pattern of the two series is not very different, partly due to the strong seasonality of the tourism series, with the biggest inflows of visitors in the summer months. Note also that, Internet searches can be seen to lead tourist inflows to some extent [6]

4 Model and regression results

The analysis presented here follows the simple procedure proposed by Choi and Varian (2009b). The strategy goes as follows.

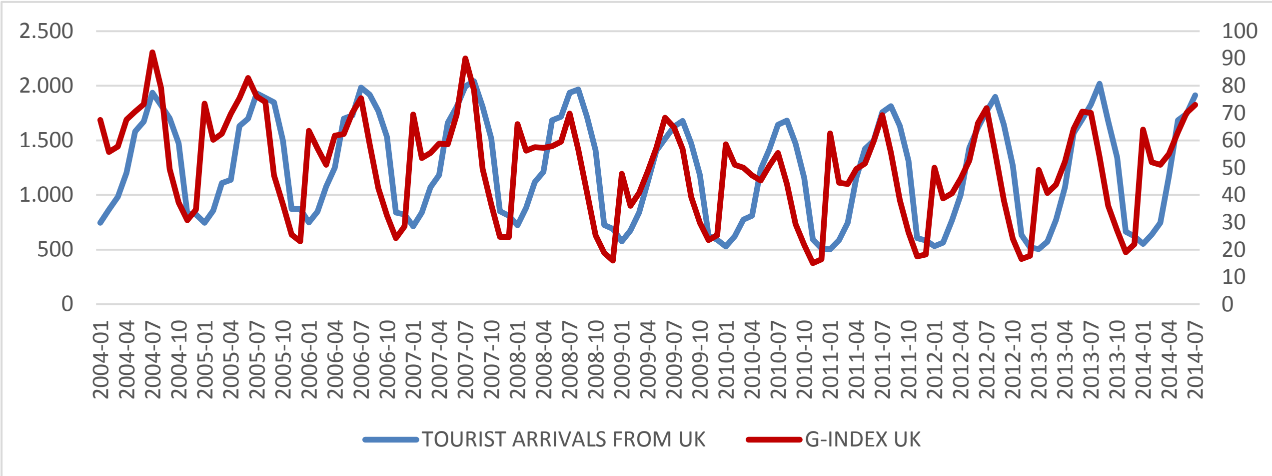
- Adjust the best possible forecasting model using the usual statistics, including the lagged endogenous variable (Model 0)
- Add the Google Trends index as an additional explanatory variable (Model 1).

- Assess the improvement in the predictions. This is typically done through the mean absolute error (MAE) of the out-of-sample predictions using a rolling window forecast.

TOURIST INFLOWS and Google Searches: By originating country **CHART 2**

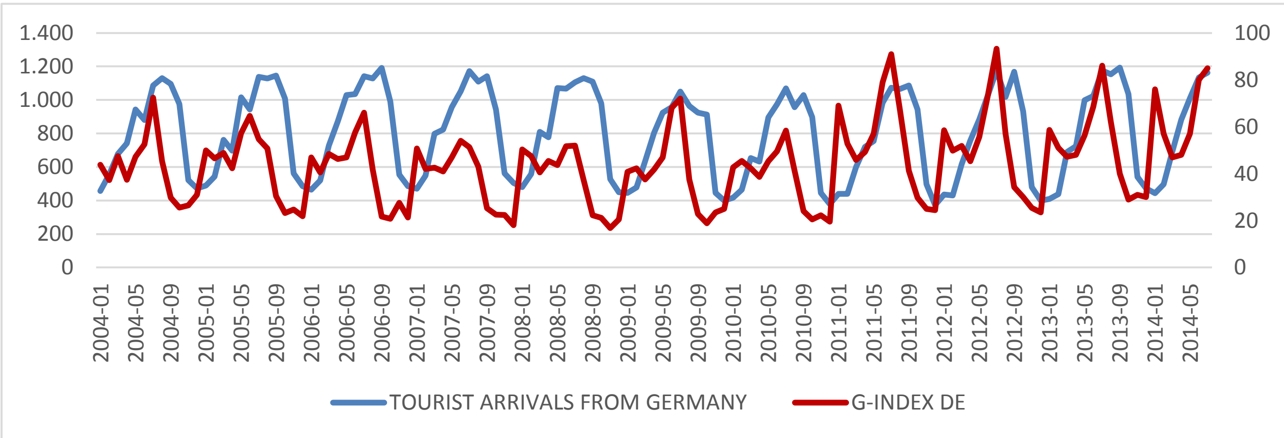
BRITAIN

Thousand People G-Index



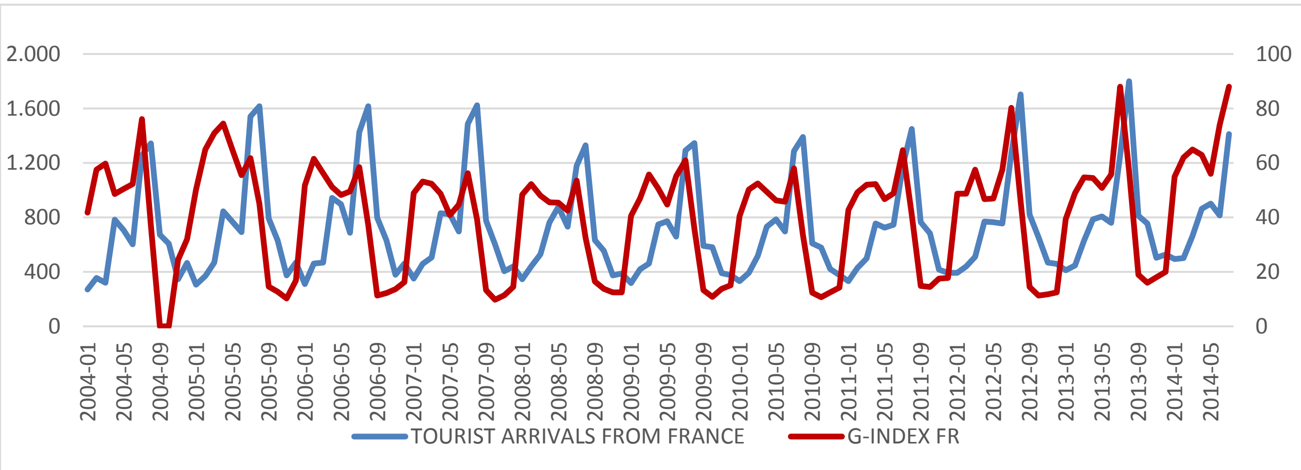
GERMANY

Thousand People G-Index



FRANCE

Thousand People G-Index



SOURCE: Instituto de Estudios Turisticos and Google Trends

Let us denote by $T_i^t = Tourist\ inflows\ from\ country\ i\ in\ month\ t$

Our baseline (model 0) is an ARIMA(p,d,q)x(P,D,Q) selected by the program TRAMO, Caporello and Maravall (2004), which best fits the data for the estimation period 2004 to 2010. TRAMO selects whether the variable T_i^t enters the model in levels or logs. The selected model includes trading day correction and Easter effects and detects and estimates outliers when required.

The selected specifications of model 0 for T_i^t are the following:

United Kingdom: $(1-B)(1-B^{12})T_{UK}^t = (1+\theta_1 B)(1+\Theta_{12} B^{12})e_t$

Germany: $(1-B)(1-B^{12})(1-\phi_1 B)(1-\Phi_{12}B^{12})lT_{DE}^t = e_t$

France: $(1-B^{12})(1-\Phi_{12}B^{12})lT_{FR}^t = e_t$

These are the benchmark models to be compared with alternative specifications (model 1) in which Google indices are included as an explanatory variable. In all instances the Google augmented models initially included the term $\sum_{k=0}^{6} \alpha_{t-k}^i G_{t-k}^i$ allowing for the coincident and lagged Google indices up to six months, the idea being that online searches might lead actual travel by up to several months.

The final augmented models, shown below, included only those lags statistically significant at 95%. Table 3 compares the two models for each country.

United Kingdom: $(1-B)(1-B^{12})T_{UK}^t = \alpha_{t-6}^{UK} G_{t-6}^{UK} + (1+\theta_1 B)(1+\Theta_{12} B^{12})e_t$

Germany: $(1-B)(1-B^{12})(1-\phi_1 B)(1-\Phi_{12}B^{12})lT_{DE}^t = \alpha_t^{DE} G_t^{DE} + e_t$

France: $(1-B^{12})(1-\Phi_{12}B^{12})lT_{FR}^t = \alpha_{t-4}^{FR} G_{t-4}^{FR} + e_t$

Forecasting models: ARIMA and G-index augmented (a) **TABLE 3**

	UK: ΔΔ12 MA(1)*MA(12) model and ΔΔ12 MA(1)*MA(12) model augmented using Google Index (-6)						
Models estimated by maximum likelihood.	θ	Θ		Standard error of residuals	AIC	BIC	Out-of-sample MSE
Est	-0,40	-0,48		51,44	769,2	8,1	0.4 E+04
SE	0,11	0,14					
	θ	Θ	**G(-6)**	Standard error of residuals	AIC	BIC	Out-of-sample MSE
Est	-0,49	-0,39	**3,11**	49,24	761,9	8,02	0.35 E+04
SE	0,10	0,13	**1,04**				
GAIN							**-12,5%**

	DE: (log) ΔΔ12 AR(1)*AR(12) model and ΔΔ12 AR(1)*AR(12) model augmented using Google Index (0)						
	Φ	Φ		Standard error of residuals	AIC	BIC	Out-of-sample MSE
Est	0,37	0,43		0,061	-188,26	-5,4	0.68E-0.2
SE	0,12	0,13					
	Φ	Φ	**G(0)**	Standard error of residuals	AIC	BIC	Out-of-sample MSE
Est	0,34	0,50	**0,0026**	0,1	-191,0	-5,4	0.58 E-02
SE	0,12	0,13	**0,0012**				
GAIN							**-14,7%**

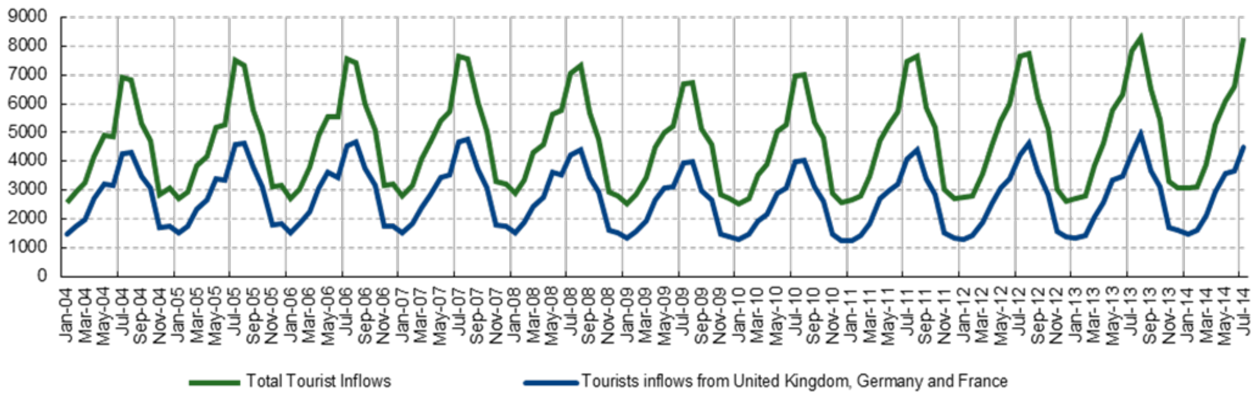
	FR: (log) Δ12 AR(12) model and Δ12 *AR(12) model augmented using Google Index (-4)						
		Φ		Standard error of residuals	AIC	BIC	Out-of-sample MSE
Est		0,14		0,058	-201,45	-5,5	0.34E-0.2
SE		0,12					
		Φ	**G(-4)**	Standard error of residuals	AIC	BIC	Out-of-sample MSE
Est		0,16	**0,0034**	0,056	-203,0	-5,44	0.19 E-02
SE		0,12	**0,0011**				
GAIN							**-44,1%**

a. Models estimated using TRAMO for the period 2006-2010. The parameters obtained in those estimates are used to make one-period-ahead up to April 2012

In all cases, model 1 appreciably improves the ARIMA model. The reduction in the MSE of the out-of-sample forecasts in the models relating to British, German and French tourist flows is 12% in the first two cases and 44% in the third, when the G-index is included. Overall these three models provide an improvement in the forecasting errors of tourist inflows from the aggregate of the three countries which, as seen in (Table 2), amount to 60% of total tourism inflows.

In order to translate these results into a workable tool which can routinely be used in the assessment and forecasting exercises for the Spanish economy we need to bridge these models to an estimation of the total tourist inflows, which is the relevant variable from a macroeconomic point of view. As seen in Chart 3, the aggregate number of travelers from the three countries is a close proxy for the total foreign tourists [7] in fact, for practical purposes for a short term forecasting exercise we can assume that their growth rates are identical.

Tourism inflows in Spain **CHART 3**



Source: Institutos de Estudios Turísticios

Denote by $\widetilde{T^t}$ the forecast obtained from the ARIMA model (Table 4) for total tourist inflows and by $\widehat{T^t}$ the forecast of total inflows derived from the Individual models based on Google searches. $\widehat{T^t}$ is calculated as: $(\dot{\widehat{T^t}}) = (\sum_{i=1}^{3} \dot{\widehat{T_i^t}})$ i.e. we assume that the growth rate of the total inflows will be the same –in the short term- than the aggregate of the three main countries.

Benchmark model for Total Tourist Inflows (Model 0)

$$(1-B)(1-B^{12})T^{t} = (1+\theta_1 B)(1+\Theta_{12} B^{12})e_t$$

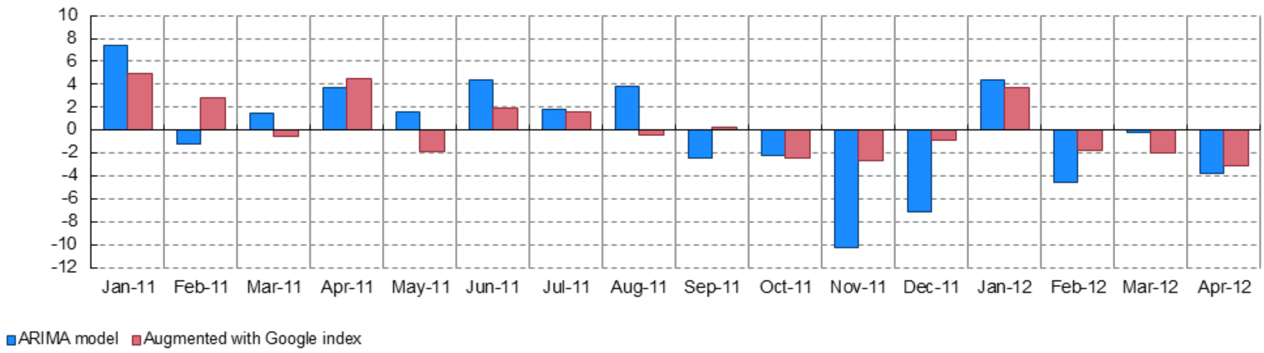
The improvement in the forecast of the total aggregate of tourists when the G-indices are added as a regressor becomes patent in Chart 4, where one-period-ahead forecasting errors are compared. The mean absolute error (MAE) for model 0 in the out-of-sample period running from January 2011 to April 2012 is 3.8%, while the MAE for the aggregate of model 1 is 2.2%.

Benchmark Model for Total Tourist Inflows **Table 4**

			ΔΔ12 MA(1)*MA(12)			
Model estimated by maximum likelihood.	θ	Θ	Standard error of residuals	AIC	BIC	Out-of-sample MSE
Est	-0,29	-0,39	156,87	927,2	10,4	3,0E+04
SE	0,11	0,13				

These G-indices appear to be performing quite well when providing short term forecasts up to 2012, beating more conventional models, and can therefore be very useful for obtaining timely economic forecasts. However, their performance worsens in the following years. This illustrates nicely that while this new tool can be a very powerful instrument for policy makers, as a valuable and timely complement for traditional statistics, further research is warranted to better understand how internet searches activities by consumers translate (or not) into actual purchases.

One Period ahead Forecasting errors for total tourism inflows **CHART 4**



5 Errors and shortcomings

In the above analysis, we showed that the inclusion of an indicator based on online searches improves our short term forecast of tourist inflows up to 2012 -reducing out of sample mean square errors by 42%- suggesting that there is valuable information in online searches that can improve the set of conventional statistics used by analysts. However, the performance of our models worsens in the following years. This reminds us of how the formerly successful use of Google search terms to identify flu outbreaks (Ginsberg at al. 2009) failed in subsequent years. As from 2009, Google Flu Trends (GFT) has been over-predicting flu. Butler (2013) reported that GFT trends was predicting more than double the proportion of doctor visits for influenza-like illness. Our exercise confirms that Google trends may provide good results now but indices and usage can change rapidly, algorithms change on a regular basis and the user base can move away on a rumour or policy or any other reason finally not translated into an actual outcome.

We can speculate on the possible causes of the worsening of model 1's performance: a change in the search terms used by potential travelers when organizing their trip [8], an increase in searches for information about holidays in Spain not translated into an increase of tourists visiting the country due to unexpected events [9]. Many alternative search terms could be explored, evaluated and compared, but in any case, more research is required in order to understand the dynamics of search terms.

To better understand how queries on vacationing in Spain might translate into actual trips might involve the disclosure of one or another type of proprietary data. Google trends only makes available indices of relative popularity instead of absolute volume of searches from a given location. Going back to the GFT example, in October 2014, Google updated the GFT prediction tool. [10], unfortunately, Google is not providing information about what exactly has changed or how Google's service works in detail including search terms used in the new version. An obvious reason is that it would give Google's competitors information about how its search engine works.

The Proprietary nature of data is not the only shortcoming. Potential sources of errors are also difficult to study. First, because Google trends data are not the result of a sampling process designed to produce valid and reliable data. They are a by-product of internet search activities that generate a lot of cases but very few variables. Google trends are case-rich but variable-poor (Prewitt 2013). This makes it impossible to study the bias and sources of errors resulting from the data generation process, or to conduct multivariate analyses, or to know why people use a specific term and/or what they do afterwards. In this sense, the above forecasting exercise, and the new version of GFT, are good examples to highlight that Google Trends are a complement, rather than a substitute, to traditional data. [11],
Among the growing number of publications using Google trends, there is a kind of a *file drawer effect* (Rosenthal 1979): an overwhelming majority of published research using Big data reports successful stories. Very little is known about their limitations and how to overcome them (Couper 2013). Only a tiny minority of papers in the field refers to case studies where the inclusion of online searches does not improve the forecasting properties of conventional models [12], e.g. Artola and Galan (2012) made the case about the relevance of the chosen benchmark model.

.

To sum up, Google trends, and Big Data in general, need deeper study and conceptualization of their sources of error and data quality. In order to develop this research line, we need better access to data and to keep an eye on traditional data. Even if, as stated by Cukier and Mayer-Schönberger (2013) , in the era of Big Data *society will need to shed its obsession for causality in exchange for simple correlations*: *not knowing why but only what,* scientist and policy makers need to know and test systematically not only what is there but also why it is there. In our case, to be able to study more in depth the causal relationship between query data and economic outcomes, and why our well performing models may break down.

6 Conclusions and future developments

Given the central role of tourism in the current account recipes a step forward should include setting a collection of appropriate key words so as to determine which are the main segments of interest for the potential travelers: ie. When looking for search terms related to vacationing in Spain they might include terms associated with a low level of expenditure (e.g. camping, cheap, etc.) or, on the contrary they may include key words associated with a higher level of expenditure (e.g. golf, sailing, *paradores*)[1]. This framework might complement the forecasting of tourist inflows and, ultimately help in projecting tourist expenditure. In addition, future research and success of Big data will depend very much on deeper access to data for researchers, and the view of Big Data as a supplement rather than a substitute of traditional data.

REFERENCES

Artola, C. and E. Galan (2012). Tracking the future of the web: constructing of leading indicators using internet searches. *Banco de España, Documentos Ocasionales.Nº1203.*http://www.bde.es/f/webbde/SES/Secciones/Publicaciones/PublicacionesSeriadas/DocumentosOcasionales/12/Fich/do1203e.pdf

Askitas, N. and Zimmermann, K. (2009). “Google Econometrics and Unemployment Forecasting”. *Applied Economics Quarterly*, 2009, 55 (2), 107-120

Askitas, N. (2014) Open Data and Data tax for the digital era https://www.linkedin.com/pulse/article/20141002074201-17164464-open-data-and-a-data-tax-for-the-digital-era?report%2Esuccess=ENi1LrS0ADyUk1yfn7ECQUR61DNkM3U7kr0x_iZnSrrr74S7_r3qBUniFRvyTSMN61SL&_mSplash=1&goback=

Butler, D. (2013). When Google got flu wrong. Nature, 494(7436), 155.

Caporello, G. and Maravall A. (2004). Program TSW: Revised Reference Manual, Occasional Paper 0408, Research Department, Banco de España.

Carrière-Swallow, Y. & Labbé, F. 2013. "Nowcasting with Google Trends in an Emerging Market," Journal of Forecasting, John Wiley & Sons, Ltd., vol. 32(4), pages 289-298, 07

[1] Paradores is a top brand hotel chain in Spain

Choi, H. , Varian H. (2009a). Predicting Initial Claims for Unemployment Benefits. July5,2009.http://static.googleusercontent.com/media/research.google.com/es//archive/papers/initialclaimsUS.pdf

Choi, H and Varian, H. (2009b), Predicting the Present with Google Trends. Google. http://static.googleusercontent.com/media/www.google.com/en//googleblogs/pdfs/google_predicting_the_present.pdf

Couper, M. (2013), "Is the sky falling? New technology, changing media, and the future of surveys", *Survey Research Methods*, Vol. 7, pp. 145–156.

Cukier, K. and Mayer-Schönberger, V. (2013). Big data: A Revolution That Will Transform How We Live, Work, and Think. Eamon Dolan/Houghton Mifflin Harcourt.

European Commission (2011). Flash EB No 328 – Survey on the attitudes of Europeans towards tourism, wave 3.

Eurostat (2011), Statistics in Focus 66/2011 "Internet use in households and by individuals in 2011

Della Penna, N. And Huang H. (2009). "Constructing Consumer Sentiment Index for U.S. Using Internet Search Patterns". Department of Economics, WP 2009-26, University of Alberta.

Ginsberg, J., Mohebbihttp://www.nature.com/nature/journal/v457/n7232/full/nature07634.html - a1 , M. , Patelhttp://www.nature.com/nature/journal/v457/n7232/full/nature07634.html - a1, R., Brammer, L., Smolinski, http://www.nature.com/nature/journal/v457/n7232/full/nature07634.html - a1 M. and Brilliant, http://www.nature.com/nature/journal/v457/n7232/full/nature07634.html - a1 L. (2009) "Detecting influenza epidemics using search engine query data". Nature Vol 457, 19 February 2009, doi:10.1038/nature07634 http://dx.doi.org/10.1038/nature07634.

Jackman, N. and Naitram S. (2013). Nowcasting Tourist Arrivals to Barbados - Just Google It! *2013 Central Bank of Barbados Annual Review Seminar, Barbados July 23-26 2013* http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2441855

Lazer, D., Kennedy, R., King, G., Vespignani, A. (2014) "The Parable of Google Flu: Traps in Big Data Analysis." Science 343 (14 March): 1203-1205. http://www.sciencemag.org/content/343/6176/1203.full

McLaren, N. and Shanbhogue, R. (2011) "Using internet search data as economic indicators," *Bank of England Quarterly Bulletin*, Bank of England, vol. 51(2), pages 134-140.

Matsumoto, A., Matsumura K., Shiraki N. (2013). Potential of Search Data in Assessment of Current Economic Conditions. BoJ Reports and Research papers. https://www.boj.or.jp/en/research/brp/ron_2013/data/ron130418a.pdf

OECD (2010), *OECD Tourism Trends and Policies 2010*, OECD Publishing, Paris. DOI: http://dx.doi.org/10.1787/tour-2010-en

Prewitt, K. (2013). The 2012 Morris Hansen lecture: Thank you Morris, et al., for Westat, et al. Journal of Official Statistics, 29(2), 223-231.

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin*, 86(3), 638-641.

Saidi, N., Scacciavillani, Ali, F. (2010) Forecasting Tourism in Dubai. Economic Note No. 8, Dubai International Financial center.

Suhoy, Tanya (2009). Query Indices and a 2008 Downturn: Israeli Data. Research Department, Bank of Israel. http://www.boi.org.il Discussion Paper No. 2009.06, July 2009.

Notes

[1] The New York Times, 11 February 2012.

[2] According to the Survey "Preferences of Europeans towards Tourism", 46% of Europeans mention Internet websites as a principal source of information for planning a holiday.

[3] It is not suggested that the second variable is a sub-set of the first one; the sizes of the two groups are simply being compared.

[4] For a detailed description of the features of Google indicators see https://support.google.com/trends/

[5] The index is robust to variations of the key words used. In fact, jointly with the index associated by Google trends to a given keyword or phrase, Google Trends provides a set of other related terms which were often queried by the same users.

[6] In addition to this lead, a rise in searches in January each year can be observed, despite the fact that in that month, and those immediately preceding it, tourist inflows are low. However, searches for holidays are relatively numerous in January, which probably reflects new plans being made as the year begins.

[7] Specifically, the correlation between total tourists and tourists from the United Kingdom, Germany and France is very close to unity, without this being merely the reflection of similar seasonal behaviour, since the correlation between year-on-year rates is likewise very high.

[8] That may be the case for returning travellers who might use search terms different from those planning a holiday to a given destination for the first time. A potential first time visitor might use the term "Spain Holidays". when searching for information about a vacation. A returning traveller who had a good experience in Mallorca on a previous trip might directly gather information on Mallorca or on a particular village or hotel on the island.

[9] This would be the case of people cancelling their trip to countries like Egypt after the Arab spring and searching for Spain as an alternative but finally not travelling because they could not find good cancelling conditions for their trip to Egypt.. In this line, it will be interesting to track how the recent turmoil involving the Ebola case detected in Spain has affected travel plans for holiday reasons.

[10] http://googleresearch.blogspot.nl/2014/10/google-flu-trends-gets-brand-new-engine.html

[11] In fact, the new GFT combines information from Google's search engine with traditional data from the Center for Disease Control and Prevention (CDC), which gathers information from 3,000 health providers Lazer et al (2014).

[12] Many of the papers take a basic Autoregressive model as a benchmark. These models are known to be quite poor in their forecasting performance.