
*Extended abtract*

# EXTENDED ABSTRACT

**Title: The Catalonian Crises through Google Searches**

**Authors and e-mail of all: Concha Artola (artola@bde.es) y Javier J. Perez (javierperez@bde.es)**

**Department: DG Economics and Statistics**

**Banco de España**

**University:**

**Subject area:** 1. Methods of spatial analysis, geographic information systems and big data

**Abstract:** In this paper we provide an application of the potential of Google Searches to track economic events in real time. We focus on the period of political turmoil starting in September 2017 in Catalonia. Our research question is the following: can the Catalan crisis be tracked by the searches done by the public on different consumption items in the Internet? We do so by focusing in two set of consumption categories: Travel to Catalonia from the main international markets (France, Germany and United Kingdom) and searches on the main consumption categories done from Catalonia and from other five big regions (Madrid, Valencia, Aragón, Andalucía and Basque Country).

The preliminary results show that the uncertainty in the political situation translated unto a decline in searches on terms associated with tourism activities in Barcelona, with one broad measure showing that searches for the term "Barcelona hotel" declined by 12%, year on year from September 2017 to January 2018. By comparison, searches for "hotel" in other comparable Spanish regions increased slightly in the same reference period. When comparing searches done from Catalonia with other regions through simple time series models, a sizeable negative residual for Catalonia is present in October 2017 –the most difficult period in the Catalan conundrum- which is not observed in other geographical areas. This is the case for some search topics associated to durable goods and Catering and Accommodation services.

## 1. Google Trends as measure of "preparatory steps to spend"

It is a well-known fact that most individuals start their consumption decision journey in the Internet. When focusing in Consumption goods, this is the case both when the final destination is online shopping and a brick & mortar shop. Among specialists, the latter case is known as ROPO or ROBO, standing for Research online, Purchase (Buy) offline. When focusing on Services, the role of internet searches can be even stronger; this is the case in Travelling services, where it is common that the entire shopping process from research to purchase happens online (which is visible by the fast decline of travel agencies), and also in Insurance and Hospitality (Hotels and Catering).

In this framework, Google Trends (GT) is a tool providing useful information on searches done by users on different topics, which might reflect "preparatory steps to spend". In this paper several GT indicators are compared for different Spanish regions around the period of political turmoil in Catalonia.

## 2. Timeline of the Catalonian crisis

The political turmoil in Catalonia started in September 2017, and the unrest reached a peak in October. We use the Dow-Jones FACTIVA database as an objective measure of the length and intensity of the unrest as seen through international newspapers. The numbers go from 30.000 "general political news" or "economic news" in normal times to 80.000 in September 2017 and 150.000 in October. In the following months public attention declined quite fast, stabilizing in 2008Q1 around 50.000 news.

## 3. Tourism in the Catalonian crisis as seen through GT

Google Indexes on Worldwide searches on the topic "Barcelona" and a breakdown for the three single countries originating the largest flows of visitors to both Catalonia and Spain. Google searches on the topic "Barcelona" from locations in the UK, Germany and France - the three countries originating the largest flows of visitors to both Catalonia and Spain- show a decline in 2017Q4, deeper in the case of France, the region originating about one quarter of visitors to Catalonia. This decline is, particularly for France and Germany quite distinctive of the search pattern for other tourist areas in Spain. These preliminary results show how the use of Google trends can be a powerful potential tool in order to get an almost real time assessment of the outlook for the tourism industry. Further research is needed in order to assess the dynamics and lags between searches and actual travels.

## 4. Consumers´ in the Catalan turmoil: How to treat Google Searches from a Regional Perspective

We match each of the consumption categories, as defined in the CLASSIFICATION OF INDIVIDUAL CONSUMPTION BY PURPOSE (COICOP), with the related Google topic. We have focused on the main items, which overall account for about half of Private Household

The first idea is to simply compare Google index for item $x$ in Catalonia and in the other big regions in Spain. If potential customers in Catalonia where delaying their consumption due to the uncertainty derived from the political situation, we´d expect this retrenchment to

translate in a lower volume of searches related to this particular consumption item relative to the trend observed in other regions. So we may expect the differential behavior of consumers from Catalonia to translate unto a unique Google trend index for topic $x$ in Catalonia, in the same way that we observed a decline of interest in Barcelona which wasn´t seen in other tourist regions in Spain. However, this simple approach might be misleading when TOTAL searches evolve differently across regions, as Google indexes do not reflect search volumes, but only measure **relative** interest for a given topic $x$ in different regions $\mathcal{R}$.

The Google Index $G_{t,I}^{\mathcal{R}}(x)$ for a topic $x$, in time t for region $\mathcal{R}$ and time lapse I measures a scaled and normalized measure of the volume of queries.

$$G_{t,I}^{\mathcal{R}}(x) = \frac{x_t^{\mathcal{R}}}{\Omega_t^{\mathcal{R}}} C_I^{\mathcal{R}} \text{ where } C_I^{\mathcal{R}} = \frac{100}{\max_{t \in I}\left(\frac{x_t^{\mathcal{R}}}{\Omega_t^{\mathcal{R}}}\right)}$$

Where: $x_t^{\mathcal{R}}$ is the number of searches on topic $x$ for time $t$ in region $\mathcal{R}$, $\Omega_t^{\mathcal{R}}$ is the number of total searches, on any topic, for time $t$ and region $\mathcal{R}$ and $C_I^{\mathcal{R}}$ is the scale term.

An increase in $G_{t,I}^{\mathcal{R}}(x)$ reflects a higher relative search volume for term $x$ with respect to all other search terms.

$$\dot{G}_{t,I}^{\mathcal{R}} = \dot{x}_t^{\mathcal{R}} - \dot{\Omega}_t^{\mathcal{R}}$$

Because it is likely that in Catalonia, during the period of political turmoil, $\Omega_t^{\mathcal{R}}$ will have increased by a higher rate than in other regions, as people in the region will be looking more intensively for news in the internet, we might observe a unique downward trend for $G_{t,I}^{\mathcal{R}}(x)$ in Catalonia which might not be related to devoting lower attention to consumption intentions on item $x$ but to an increase in overall queries in the Web.

Given that we don´t know anything about the size of $\Omega_t^{\mathcal{R}}$ we must device a method to overcome this problem.

## Getting rid of normalization

Compare $G_{t,I}^{\mathcal{R}}(x)$ with $G_{t,I}^{\mathcal{R}}(\mathcal{W})$ where $\mathcal{W}$ is a search term selected as a reference, from which we have external information about its popularity.

The ratio of the two indexes give us the actual volume of searches on $x_d$ as a fraction of searches on the reference term $\mathcal{W}_t^{\mathcal{R}}$

$$\frac{G_{t,I}^{\mathcal{R}}(x)}{G_{t,I}^{\mathcal{R}}(\mathcal{W})} = \frac{x_t^{\mathcal{R}}}{\mathcal{W}_t^{\mathcal{R}}}$$

Selecting a suitable reference term

Finally, we selected Football, $\mathcal{F}$, as a reference term, given that it shows some desirable properties: it has no zeroes at the beginning of the sample and it is very stable across regions and along time.

Modelling the series $x_t^{\mathcal{R}}/\mathcal{F}_t^{\mathcal{R}}$

We model the series $x_t^{\mathcal{R}}/\mathcal{F}_t^{\mathcal{R}}$ for the biggest six Spanish regions + SPAIN and 14 consumption items $x$ (98 time series) that we model using the automatic option available in TRAMO_SEATS for Windows.

We estimate the automatically selected ARIMA models for the monthly series $x_t^{\mathcal{R}}/\mathcal{F}_t^{\mathcal{R}}$ for the time lapse 2004m1-2017m1 and compute the one_ period_ ahead errors for the last year (2017m2-2018m1).

## 5. Results and Discussion

Comparing Catalonia with Madrid and Spain as a whole

- In some cases, namely Apparel and Cars, Mobile Phone, Computers, among durable goods there is sizeable negative residual for Catalonia in October 2017, which is not observed in the other two geographies. This is also the case in Theaters (but not in other cultural services)

- These negative spikes in October - the most difficult period in the Catalan conundrum- might suggest the political turmoil had a negative impact in consumption prospects, which were not permanent.

In order to get a cleaner view of the possible incidence of Catalonian turmoil on Consumption we go through the following exercise: We estimate automatic ARIMA models for 98 series for the period 2004m1- 2015m12. We then filter the one_period_ahead residuals for the 14 consumption items in Catalonia on the five main regions (Andalucia, Aragón, Madrid, Pais Vasco and Comunidad Valenciana) ARIMA residuals.

The results show the presence of a negative spike in October 2017 which is significant only in two consumption categories: Theater and Restaurants.

GT as a timely indicator, useful in the presence of events, has some value for policy makers and analysts


The trick of get rid of normalization (first applied in a different context by Askitas & Zimmerman) can be a useful tool when dealing with regional studies.

**Keywords: Keywords:** *Google Trends, International Tourism, Catalonia, Private Consumption*
**JEL codes: C81, E29, L86, Z30**