

Tracking the future on the web: using Google Trends to improve tourist inflows and auto sales forecasting in Spain

Concha Artola

Economic Analysis and Forecasting

14th IWH-CIREQ Macroeconometric Workshop
Halle (Saale), December 2 2013

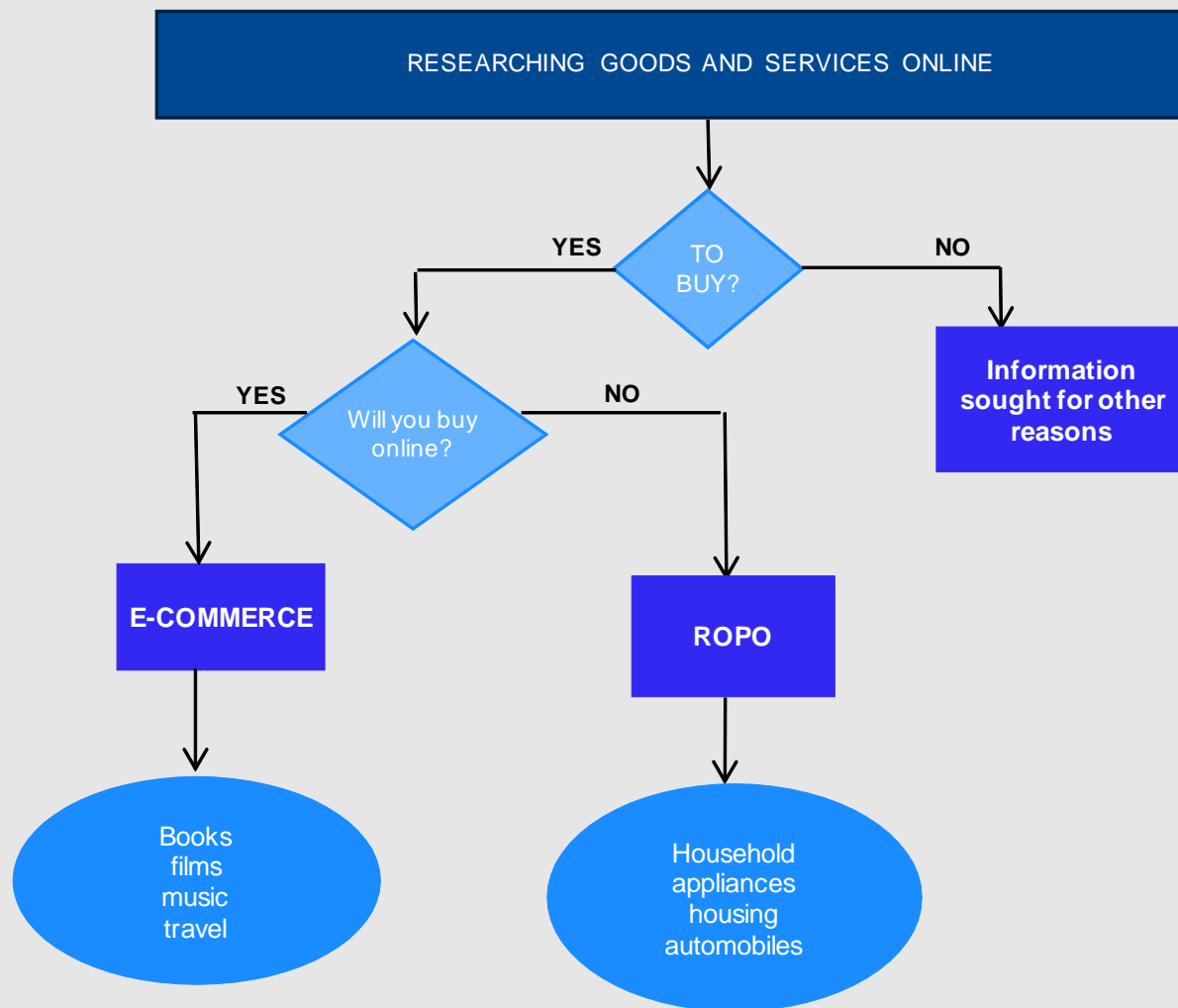


- **Searches and E-Commerce: What's better for constructing economic indicators?**
- **Economic indicators based on web searches.**
 - *How does Google Insights for Search work?*
 - *In the toolkit of Central Banks*
- **Forecasting tourist inflows to Spain using Google searches**
 - *Results*
 - *Warnings*
- **Nowcasting auto sales**



- **Searches and E-Commerce: What's better for constructing economic indicators?**
- Economic indicators based on web searches.
 - *How does Google Insights for Search work?*
 - *In the toolkit of Central Banks*
- Forecasting tourist inflows to Spain using Google searches
 - *Results*
 - *Warnings*
- Nowcasting auto sales

Searches and E-Commerce: What's better for constructing economic indicators?

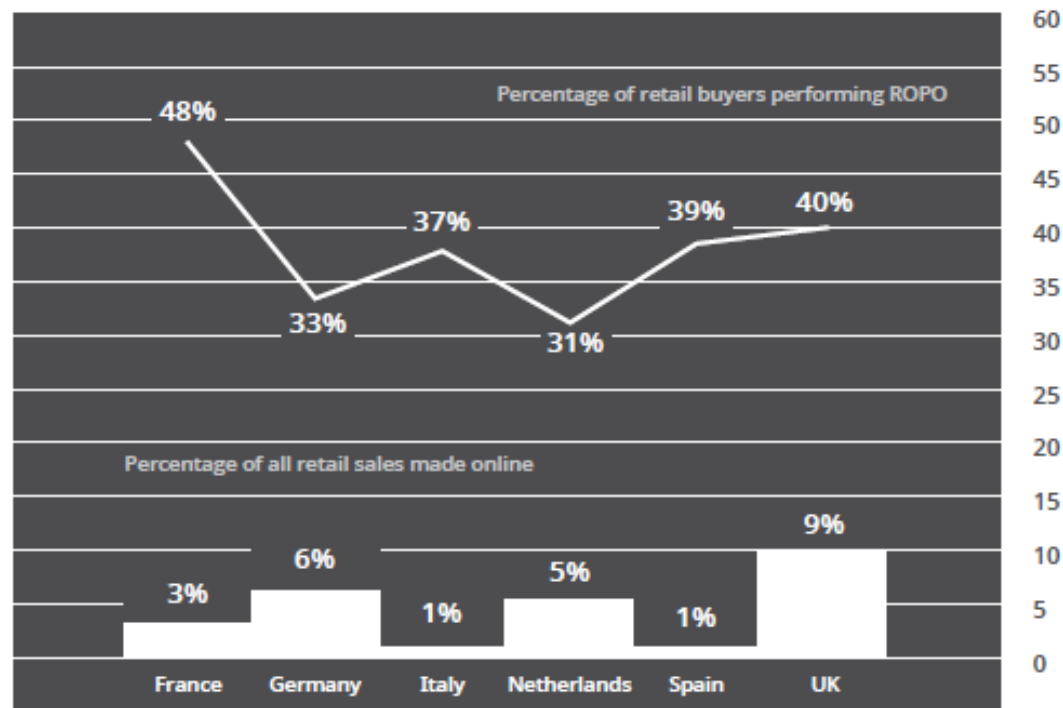


Searches and E-Commerce: What's better for constructing economic indicators?



1. The most common activity carried out by internet users –after email- is **searching** for info on goods and services. 80% of internet users (Eurostat)
2. In 2011, 58% of users have **shopped** on line at least once. Huge differences accross countries. 30% purchased trips or made hotel reservations

Fig. 1 Today's buyer is more informed than ever





- Constructing economic indicators based on Web activities
 - *Searches & E-commerce*
- **Economic indicators based on web searches.**
 - *How does Google Insights for Search work?*
 - *In the toolkit of Central Banks*
- Forecasting tourist inflows to Spain using Google searches
 - *Results*
 - *Warnings*
 - *Future developments*
- Conclusions

Economic indicators based on internet activity

How Google Trends works?



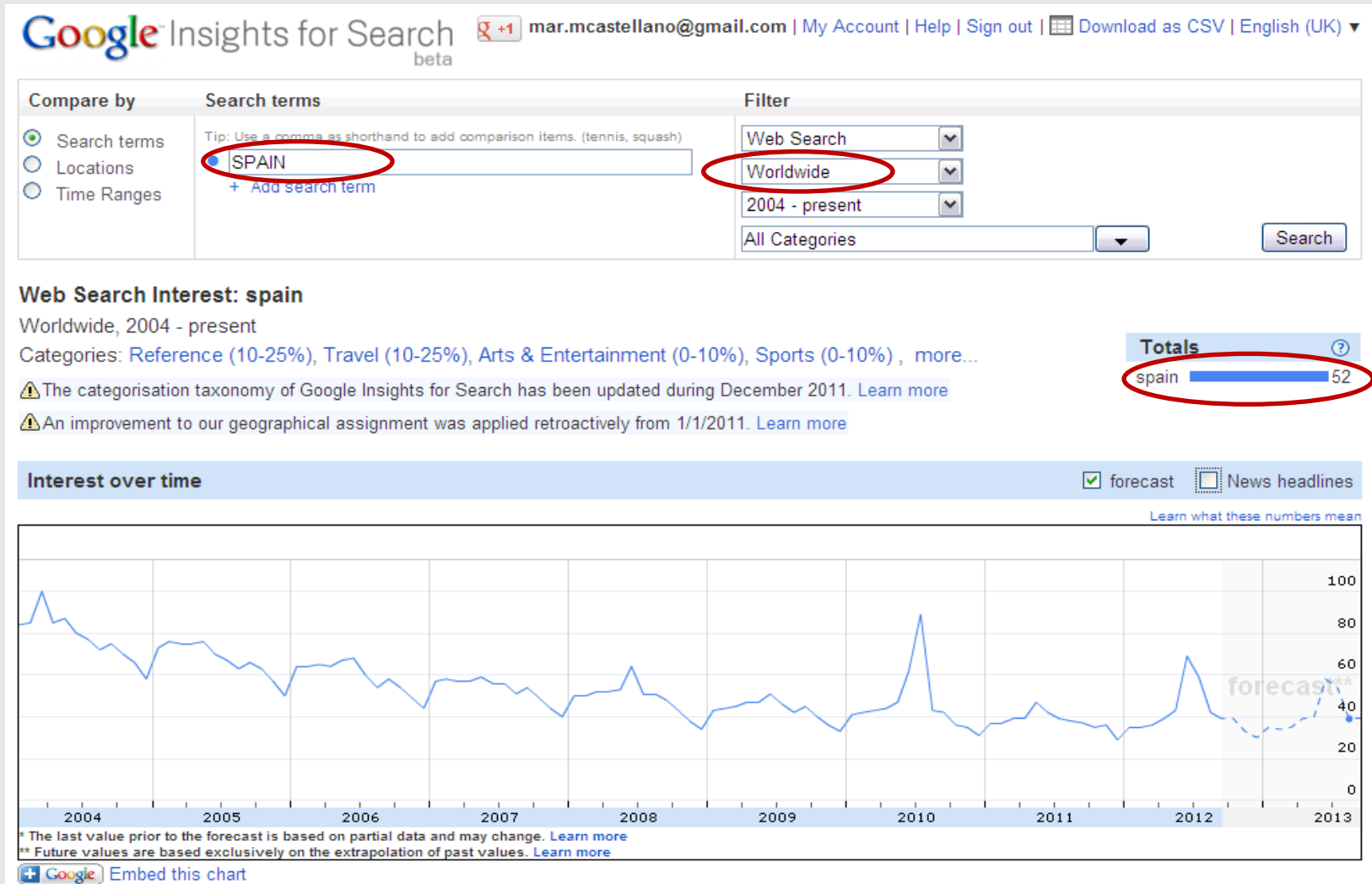
Given the prevalence of online searches versus online shopping, it seems sensible focusing in the former when constructing economic indicators of activity.

This line of research was pioneered by Choi & Varian in their paper “Predicting the Present with Google Trends” (2009) and relies on a statistical tool designed by Google which allows to compile and analyze the zillion searches made by Internet users.

How the Google tool works?

Economic indicators based on internet activity

How Google Insights for Search works?



Google Insights for Search

Economic indicators based on internet activity

How Google trends works?



Google Insights for Search beta mar.mcastellano@gmail.com | [My Account](#) | [Help](#) | [Sign out](#) | [Download as CSV](#) | [English \(UK\)](#) ▼

Compare by	Search terms	Filter
<input checked="" type="radio"/> Search terms	Tip: Use quotation marks to match an exact phrase ("table tennis"). <input type="text" value="SPAIN"/> + Add search term	<input type="text" value="Web Search"/> ▼
<input type="radio"/> Locations		<input type="text" value="Worldwide"/> ▼
<input type="radio"/> Time Ranges		<input type="text" value="2004 - present"/> ▼
		<input type="text" value="Travel"/> ▼
		<input type="button" value="Search"/>

Web Search Interest: spain

Worldwide, 2004 - present

[All Categories](#) > [Travel](#)

Subcategories [Tourist Destinations \(25-50%\)](#), [Hotels & Accommodations \(10-25%\)](#), [more...](#)

⚠ The categorisation taxonomy of Google Insights for Search has been updated during December 2011. [Learn more](#)

⚠ An improvement to our geographical assignment was applied retroactively from 1/1/2011. [Learn more](#)

⚠ Fewer than 25% of searches containing your search terms belong to the Travel category. [Learn more](#)



Interest over time

☒ forecast ☐ News headlines

☒ Interest level ☐ Growth relative to the Travel category

[Learn what these numbers mean](#)



[Embed this chart](#)

Google Insights for Search

Economic indicators based on internet activity

How Google Trends works?



Select a search term. (e.g. SPAIN)

We control for the **geographical area** where searches are generated
the **time period** (starting at 2004)
and the **category** (here all categories).

Results and meaning:

- The chart represents relative interest in search terms including the word SPAIN.
- Strong Seasonality
- The number in the bar (50%) DOESN'T Mean that 50% of overall searches include the term SPAIN:
- For a sensible reading one must know that Google indexes are both scaled and normalized. This means:
 - Data are normalized meaning we don't know the absolute figures for search volume, only the relative frequency. In top of that, data are scaled in a range [0,100], 0 meaning non enough searches, 100 is the top interest
 - Besides these very salient facts, the seasonality of the time series is noteworthy. This is better seen in the second chart where only searches about Spain lying into the TRAVEL category are included

Caution should be exercised in some aspects.

Economic indicators based on internet activity

In the toolkit of Central Banks



Bank of Israel: launched as soon as 2009 an index based on Google searches which serves as an indicator of demand in the economy, whose results are regularly included in the press release following the Monetary Committee.

Bank of England: tracks internet searches in specific fields, namely unemployment trends and housing market developments (QB 2011Q2)

New York Fed: applied Google Insights for search to the mortgage markets, finding a sizeable improvement in the outlook for refinancing applications filled when Google searches on mortgage refinancing are included in the model.

Bank of Australia: applies Google searches to two specific fields, tracking online shopping and unemployment developments (QB June 2012)

Bank of Spain: tracks tourist inflows. DO published in March 2012 and box in QB July 2012. Other projects ongoing.



- Introduction: Internet economy generates Big Data.
- Penetration and use of the Internet in Europe
 - *What activities are carried out by users?*
 - *E-commerce*
 - *ROPO: Research Online-Purchase Offline*
- Economic indicators based on internet activity.
 - *How does Google Insights for Search work?*
 - *In the toolkit of Central Banks*
- **Forecasting tourist inflows to Spain using Google searches**
 - ***Results***
 - ***Warnings***
 - ***Future developments***
- **Conclusions**



Can Google beat the performance of conventional short-term forecasting models?

- **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.**

Standard procedure as seen in
Choi , Varian & others

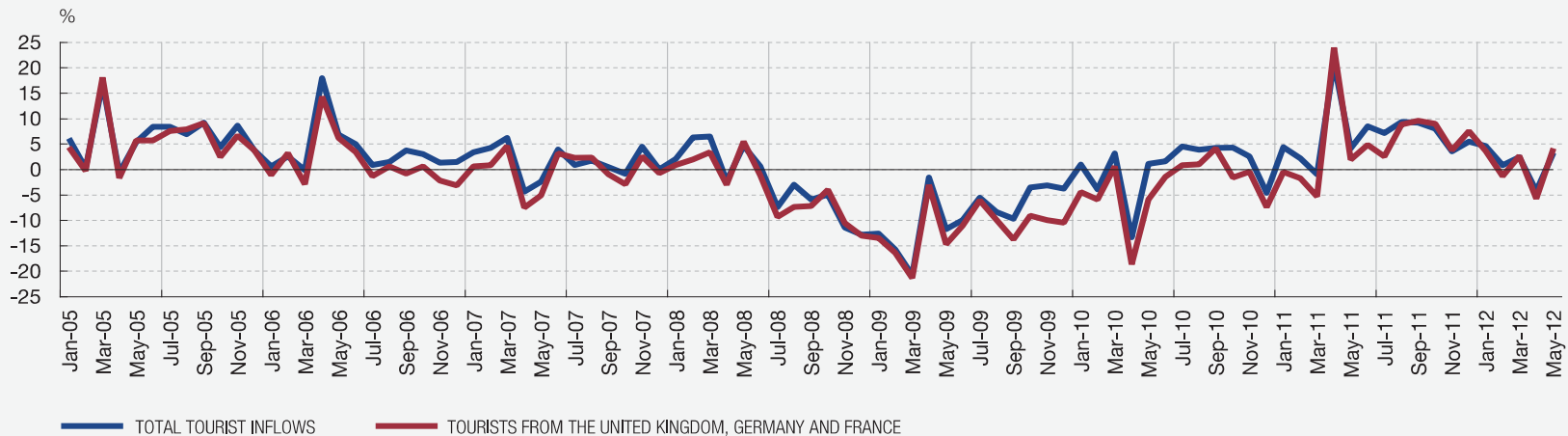
Does Model 1 forecast better than Model 0?

Forecasting tourist inflows to Spain using Google searches

Results

The indicator constructed only considers tourists from the UK, Germany and France, the three main clients of the Spanish tourism industry. Overall, these three countries account for 60% of the foreign tourists in Spain, and their tourism pattern is very similar to that of the aggregate of total foreign tourists

1 TOURIST INFLOWS INTO SPAIN
(year-on-year rates)



It looks like good enough to focusing on travellers from these three countries

Forecasting tourist inflows to Spain using Google searches

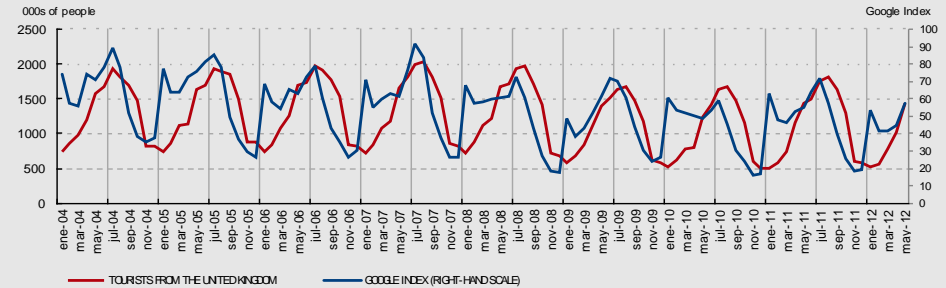
Results

United Kingdom: 1 to 2 months

- Correlation (0): 0.51
- Correlation (-1): 0.75
- Correlation (-2): 0.78

TOURISTS INFLOWS INTO SPAIN

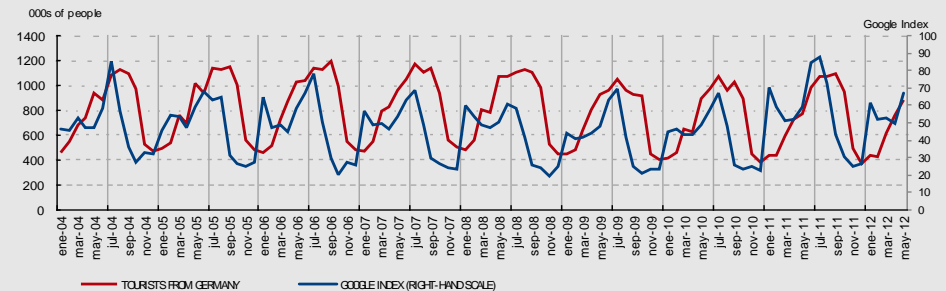
TOURISTS FROM THE UNITED KINGDOM



Germany: 2 months

- Correlation (0): 0.30
- Correlation (-2): 0.80

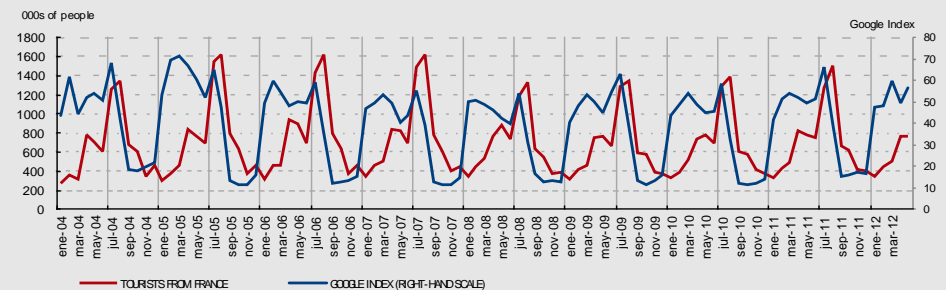
TOURISTS FROM GERMANY



France: 1 to 3 months

- Correlation (0): 0.26
- Correlation (-1): 0.57
- Correlation (-2): 0.57
- Correlation (-3): 0.56

TOURISTS FROM FRANCE



SOURCES: Instituto de Estudios Turísticos y Google Insights for Search

Forecasting tourist inflows to Spain using Google searches

Results

FORECASTING MODELS (a) ARIMA models of tourist inflows

Germany: Model (log) $\Delta\Delta_{12}$ AR(1)*AR(12) and using Google (lag=0)								
Models estimated by maximum likelihood	Φ	Φ		STD	AIC	BIC	out-of-sample MSE	
	Est	0.37	0.43		0.1	-188.3	-5.4	0.68 E-0.2
	SE	0.12	0.13					
	Φ	Φ	G	STD	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%	

Source: Bank of Spain

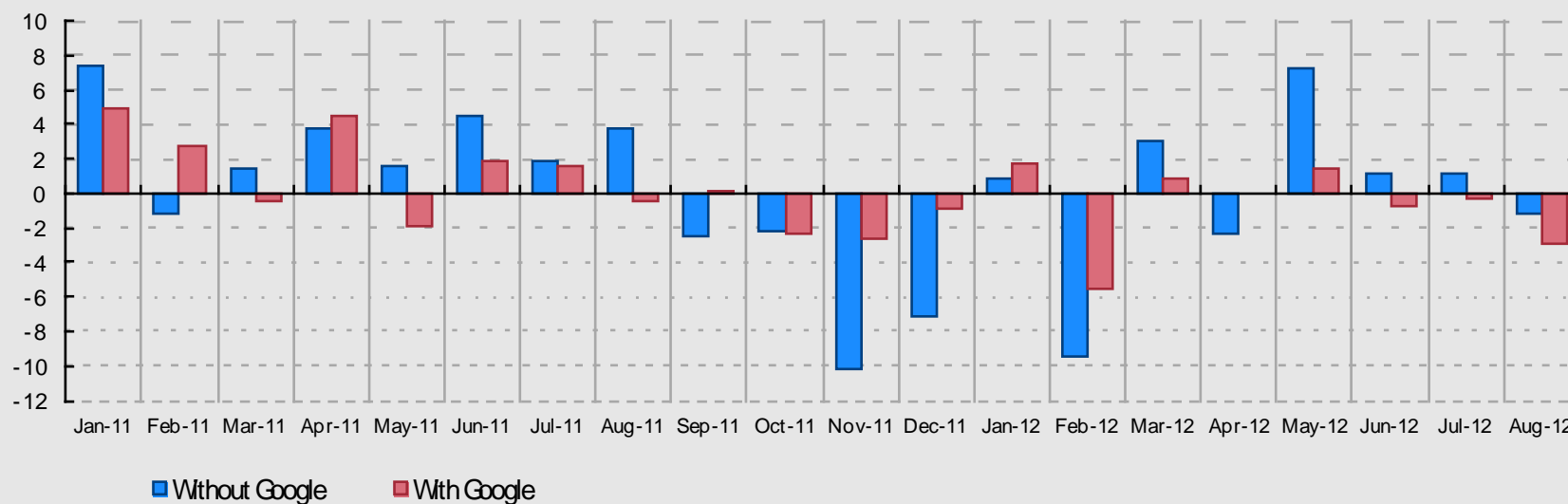
a. The models have been estimated in TRAMO-SEATS. Sample period: 2004 to 2010

Forecasting tourist inflows to Spain using Google searches

Results



ONE-PERIOD AHEAD FORECASTING ERRORS



Model 0: MAE = 3.8%

Model 1: MAE = 2.2%

Forecasting tourist inflows to Spain using Google searches



Results

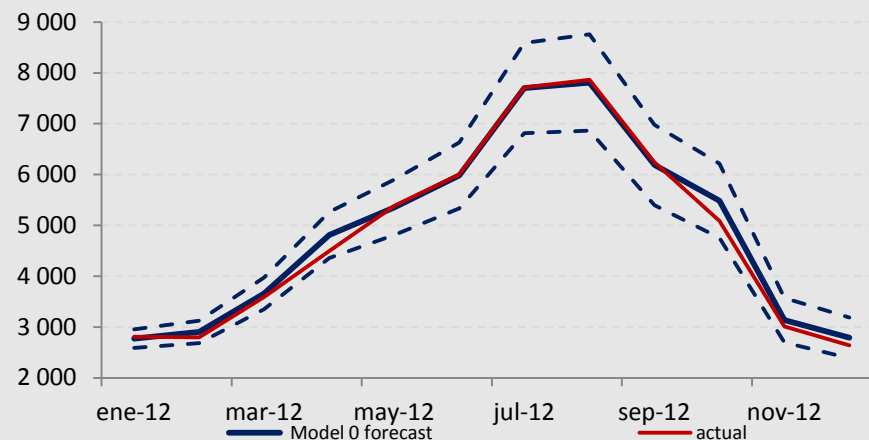
Forecast of Tourist inflows to Spain

Origin December 2011

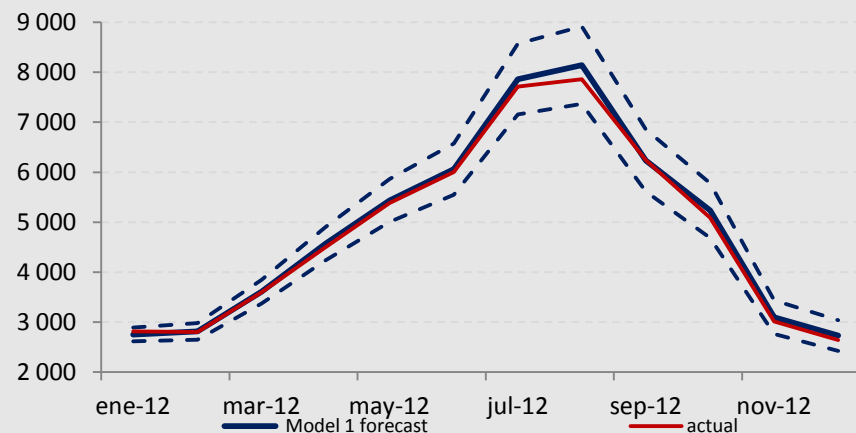
	Model 0	Model 1
Forecast for 2012	58,594	58,538
Upper i confidence	65,150	63,652
Lower i. confidence	52,038	53,423
Magnitude of ic	11.2%	8.7%
		-21.9%
Observed 2012	57,636	
error	1.66%	1.57%
		-5.9%

Forecasts with origin Dec 2011

Model 0



Model 1



OUTLINE



- Searches and E-Commerce: What's better for constructing economic indicators?
- Economic indicators based on web searches.
 - *How does Google Insights for Search work?*
 - *In the toolkit of Central Banks*
- Forecasting tourist inflows to Spain using Google searches
 - *Results*
 - *Warnings*
- **Nowcasting auto sales**



Nowcasting Auto sales

Forecasting auto sales using Google searches

1. searching terms: “compra coche” “coche nuevo” ...
2. By brand name.

Both failed: looks difficult to determine a procedure which adequately capture the intention to buy a car. **What are the key words used by potential shoppers?**

Possible explanation: CONSTANT CONSIDERATION (Google study for US consumers):

63% of final shoppers started their online search looking for a particular brand or model up to six months before the purchase.....but only 20% of shoppers end up buying the car they were initially interested in.

Nowcasting auto sales using Google searches

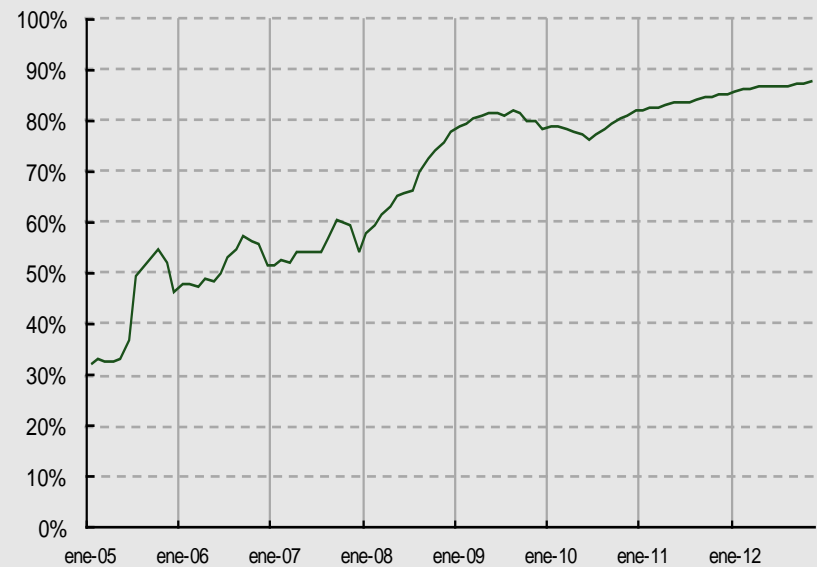
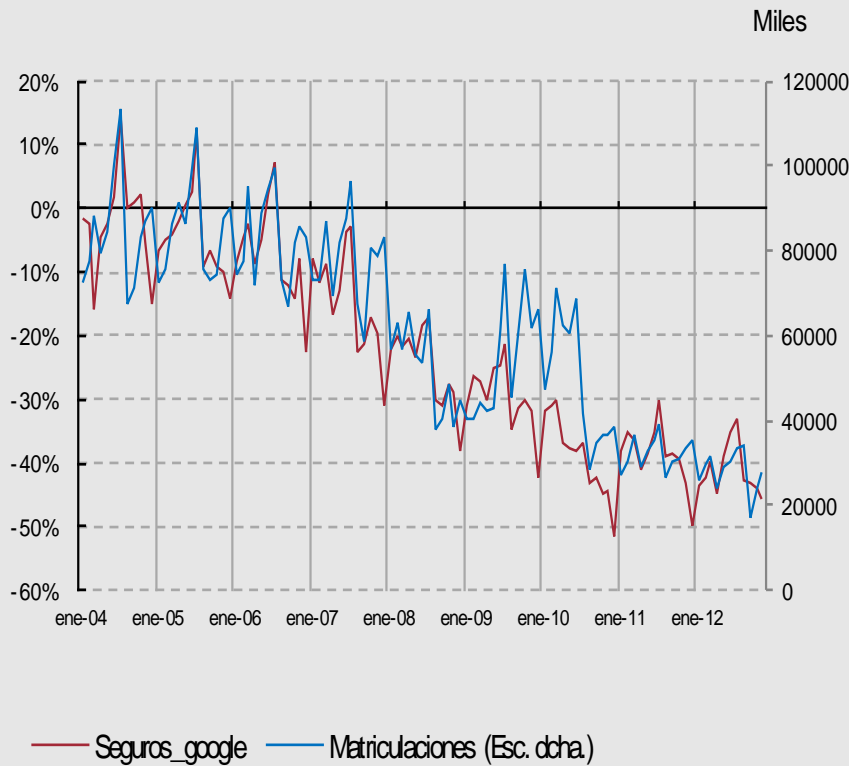
Simultaneous indicator based on Insurance



Matriculaciones de vehículos particulares

Correlación (rolling)

(a) Correlación desde enero de 2004 hasta t



(a) Correlación desde enero de 2004 hasta t

Forecasting auto sales using Google searches

Simultaneous indicator based on Insurance

Results

MODELOS DE PREVISIÓN: SIMPLES Y AUMENTADOS CON GOOGLE (a)
Modelos ARIMA de matriculaciones de automoviles particulares

CUADRO 1

Modelo $\Delta\Delta_{12}$ MA(12) y aumentado con Google (0 retardos)

Modelos estimados por Máxima Verosimilitud.	θ		Error estandar de los residuos	AIC	BIC	EAM fuera de la muestra	Outliers
Est	-0,78		0,070	-189,0	-5,1	0,0042	07/2010 (LS), 01/2008 (LS)
SE	0,07						
	θ	G	Error estandar de los residuos	AIC	BIC	EAM fuera de la muestra	
Est	-0,65	0,40	0,066	-202,3	-5,2	0,0033	07/2010 (LS), 01/2008 (LS), 06/2009 (LS)
SE	0,08	0,17					
Ganancia						-22,4%	

(a) Los modelos se han



CONCLUSIONS

The use of Big data for Economic Analysis is still at its infancy

“ Your Billion Google Searches Now a Central Bank Tool”
(Bloomberg, August 14th 2012) includes different views on the new tool. Mostly positive, although some more skeptical.

Lucrezia Reichlin: “ Google is sexy and something may come of it, but more research is needed”

San Francisco Fed John Williams referred to what he sees as a “data revolution”: “an enormous amount of information that will better help us understand in very real time what’s going on”

Indeed further research is guaranteed as suggested by Ms Reichlin, but the field is really promising: In **E. Brynjolfsson (MIT)** words: When central bankers were looking at traditional data, they were essentially looking out the rear-view mirror”



THANKS FOR YOUR ATTENTION

BANCO DE **ESPAÑA**
Eurosistema

DG ECONOMICS, STATISTICS AND RESEARCH