

CONSTRUCTION OF ECONOMIC INDICATORS USING INTERNET SEARCHES

Mioara POPESCU

Academy of Economic Studies, Bucharest

E-mail: mio.popescu@yahoo.com

Abstract

The volume of online data searches can be used as indicators of economic analysis and forecasting. This paper reviews some of the applications that use the large data sets provided by the Internet user searches and presents a very specific case for Romanian economy. These data provide some additional information relative to existing surveys and with further development, internet search data could become an important tool for analysis and prediction.

Keywords: *nowcasting, economic indicators, forecasting, big data*

JEL Classification: C15, C40, L48

Introduction

The abundance of information in terms of both volume and area of interest is giving a numerous new sources of data which can be used in a various academic, business and government area. The “internet of things” and “big data” are two of the most-talked-about technology topics nowadays. The data about users' web activity is available very quickly and the traces left by users' online activity can be used in the statistics production. For example, when confronted with an illness, users search for information on specialised health care websites and post about their mood in their Facebook page or Twitter. These data is automatically saved in big databases supporting the web services or the log files of the web servers.

There is already some experience in using this web activity data to predict statistics socio-economic indicators, like unemployment, flu incidence, real estate market activity and travel, migration flows and automotive sales.

The present study follows the line of research made by Choi and Varian, based on "Google Trends", a statistical tool designed by Google which allows the Internet users to compare the volume of searches between two or more terms. The data is made available in aggregated form by the Google web service.

This article reports a relatively easy way to combine some data on web search activity of users to improve the accuracy of simple predictive models, in the case of tourist flows.

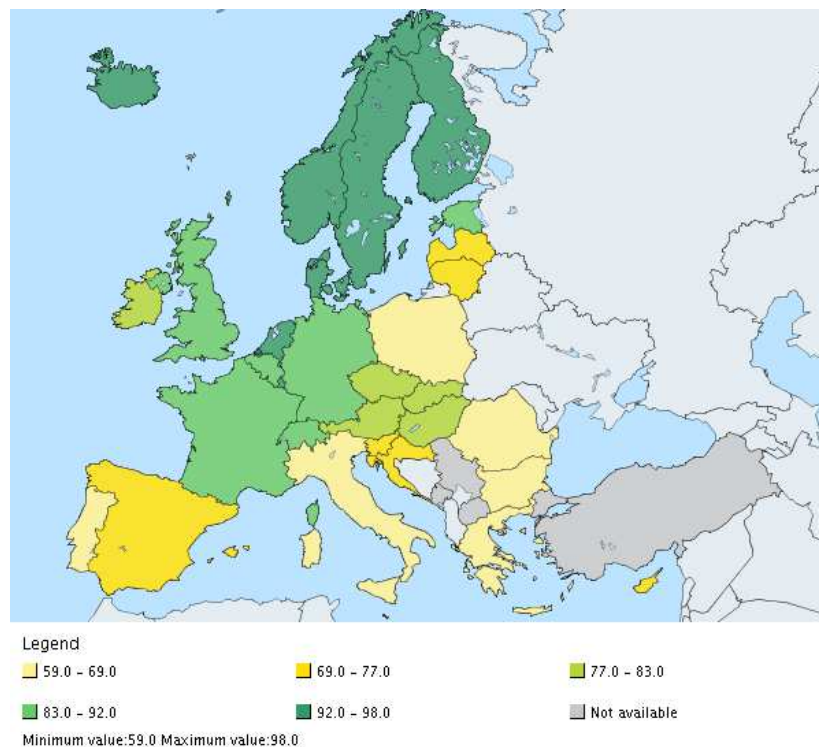
The paper is organized as follows: Section 2 summarizes the penetration and use of the Internet in Europe; Section 3 describes a model application to the Romanian economy: forecasting the Italian tourist inflows; Section 4 outlines some conclusions and proposes follow-up projects.

European Internet usage

The use of the Internet offers a large amount of information, benefits and daily activities to everyone. More than 3 Zetabytes of data exist in the digital universe today and this number has increased exponentially in recent years. The Internet has become a tool for avoiding the hassles of the bank and making the transactions quickly and safely. It also offers a powerful source for shopping, socializing and expanded the availabilities of working from home.

A recent study conducted by the European Statistical Office (Eurostat) on Internet access and use by the European households shows that almost four out of five individuals (78%) in the European Union (EU) used the internet at least once in the three months prior to the survey. The Digital Agenda target of 75% of the population using the internet regularly in 2015 was reached in 2014. The proportion of internet users who go online on a daily basis was high in all Member States and in Iceland, Norway and Switzerland. Just under two thirds of all EU citizens (65%) used the internet every day or almost every day.

Fig. no. 1. *Internet use by individuals (% of individuals aged 16 to 74 – 2014)*



Source: Eurostat.

The proportion of individuals living in the EU who have never used the internet dropped to 18% in 2014, and is just three percentage points above the Digital Agenda 2015 target of 15%. A significant “digital divide” remains, with

large differences still seen between the rates of non-use in individual countries. The proportion of the population with no experience of using the internet was highest in Romania (39%), Bulgaria (37%) and Greece (33%) and lowest in Denmark (3%), Luxembourg (4%) and the Netherlands (5%).

An application to the Romanian economy: tourist inflows

The survey on preferences of Europeans towards tourism conducted in 2014 by the European Commission (EC) shows that recommendations from friends, colleagues or relatives are considered the most important source of information when it comes to planning holidays, with the Internet second in importance. However, the Internet is the most common method for organising a holiday; it was used by two-thirds of people when arranging a holiday, an increase from 2013.

Tourism is one of the most online search-intensive activities, both for obtaining detailed information on destinations and planning holidays, and for the online purchase of touristic packages. The use of the Internet to organise holidays varies greatly at a national level, from 27% in the Former Yugoslav Republic of Macedonia to 84% in Ireland.

World Travel and Tourism Council, ranks Romania 4th amongst the tourism countries that has grown rapidly, with an annual growth of tourism potential of 8% from 2007 to 2016.

Table no. 1. **Arrivals at tourist accommodation establishments in Romania by country of residence of the tourist**

	2007	2008	2009	2010	2011	2012	2013
Italy	436005	426624	388163	435616	488244	508132	500790
Bulgaria	157765	193185	243771	260318	333974	336135	354508
Austria	227797	299859	255771	242151	254950	244000	35371
Spain	153199	196592	180764	224675	255990	234254	27504
Hungary	200681	245521	204436	204844	207836	217631	13876

Source: Eurostat.

The main clients of the Romanian tourist industry are Italy, Bulgaria, Austria, Spain and Hungary. According to data published by the Statistical Office of the EU, since the financial crisis has affected tourism in 2007, tourist inflows have fallen by 1.4% compared to 2010 and increased by 1.6% over the period.

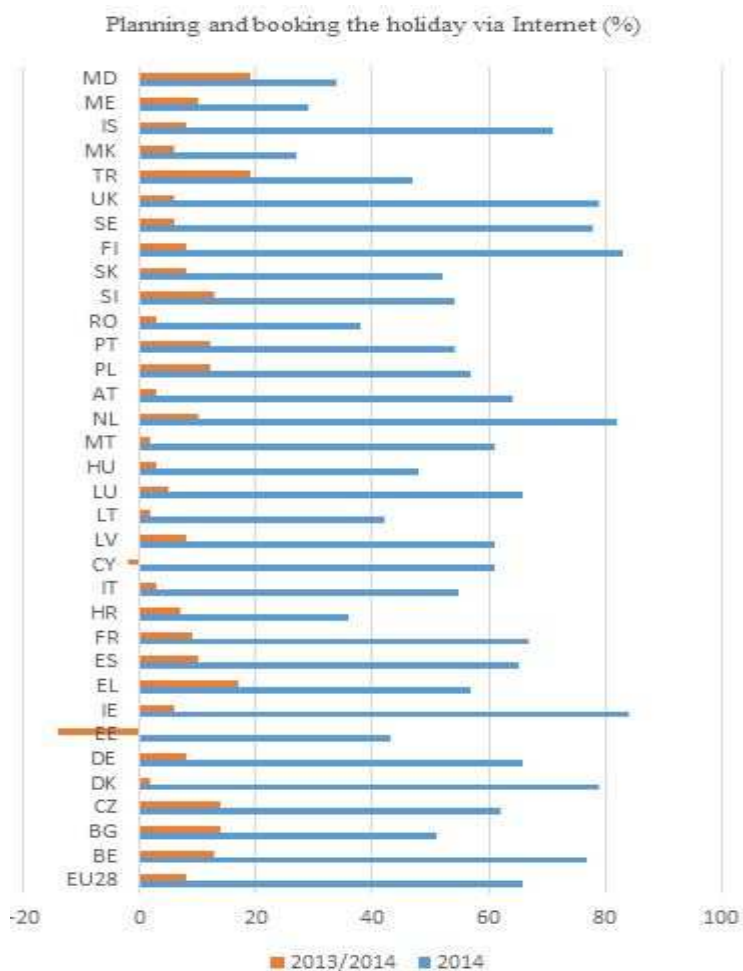
Applying Google Trends to tourist flows

According to the EC survey on preference of European towards tourism, Internet websites are the most important sources of information for respondents in seven countries, led by Finland (65%), the Netherlands (63%) and Luxembourg (58%). In Denmark, it is equal in importance with personal recommendations (56%). At the other end of the scale, less than a quarter of respondents in the

Former Yugoslav Republic of Macedonia (18%), Moldova (19%) and Romania (20%) consider Internet websites an important source.

Respondents aged 25-39 (77%) are the most likely to have booked a holiday via online services, while those aged 55 and over (51%) are the least likely to have done so. The Internet (66%) was by far the most common way of organising a holiday and its use is constantly growing (+8% since January 2014). It is particularly popular for booking short holidays.

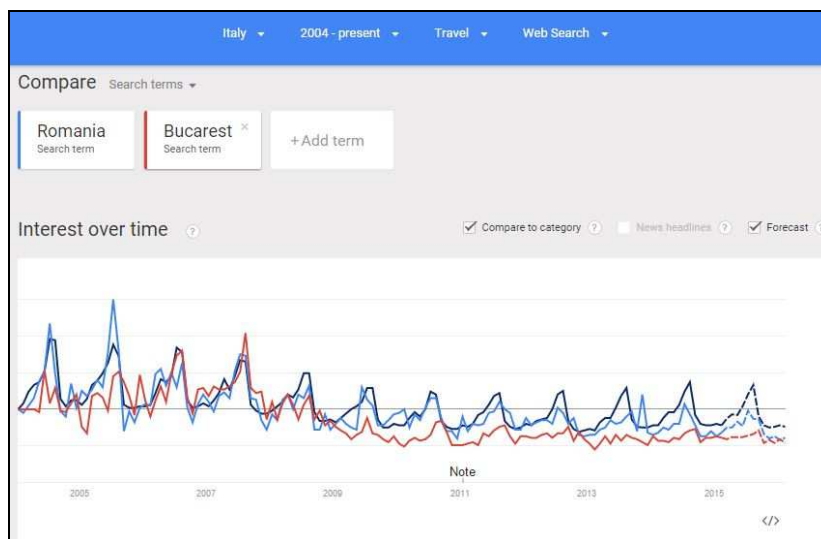
Fig. no. 2. *Planning and booking the holiday via Internet (%).
EC survey on preference of European towards tourism*



Italy is the main customer of the Romanian tourist industry and most of the Italians are using the Internet for organising holiday. Italy is ideal to use as a model in the construction of leading indicators of tourist flows.

The first point is searches made from Italy for the terms “Romania” and “Bucharest”. The interest of Italian citizens has fallen after the financial crisis in 2007. From 2009 until January 2015 the series are not very different.

Fig. no. 3. Google Trends (index) January 2004 – January 2015 and the forecast for 2015



The correlation between the two series reflects the seasonal component. The biggest inflow of Italian tourists is in July and August and reaches the minimum level in December and February. Next tables outline the results of the estimates made. Table 2 compares an ARIMA model constructed in R, which shows good forecasting properties compared with the Google index series. The mean of the forecasting errors is about 19%.

Table no. 2. MA(1)*MA(12) model and MA(1)*MA(12) model augmented using Google Index

Models estimated by maximum likelihood	θ	Θ		Standard error of residuals	AIC	BIC	Out-of-sample MSE
East	-0.11	-0.38		52.3	511.4	9.1	0.5231E+04
SE	0.15	0.20					
	θ	Θ	G	Standard error of residuals	AIC	BIC	Out-of-sample MSE
East	-0.18	-0.31	11.40	49.8	504.8	9.3	0.3991E+04
SE	0.17	0.19	3.98				

By comparing with a second model ARIMA, where the regular component is set to zero, the percentage decreased from 19% to 13%.

Table no. 3. **MA(12) model and MA(12) model augmented using Google Index**

Models estimated by maximum likelihood	θ	Θ		Standard error of residuals	AIC	BIC	Out-of-sample MSE
East	0.00	-0.35		52.1	510.3	9.0	0.3568E+04
SE		0.12					
	θ	Θ	G	Standard error of residuals	AIC	BIC	Out-of-sample MSE
East	0.00	-0.18	12.40	45.1	496.3	7.9	0.3791E+04
SE		0.13	3.51				

Briefly, the improvement in forecasting compared with conventional models depends on which ARIMA model is used as a benchmark. The information contained in Google searches could be a value for forecasts and the power of the indicators, which provide information in real-time, has to be put in the appropriate context.

Conclusions and future developments

In this paper we have analysed the field that use the big amount of information provided by Internet user searches. This valuable information, rich in volume and obtained in real time, proves to be feasible to analysts by using the statistical tools such as “Google Trends”, which allows trends in different areas of interest to be categorized and evaluated. Recent papers have concentrated predominantly on the labour market, the housing market and travel and consumer confidence.

This paper has presented a very specific application for the Romanian economy, specifically Italian tourists (the Romanian tourist industry's main consumers) visiting Romania. The amelioration in forecasting provided by the short-term models which include the Google-indicator depends on the model taken as a point of reference. In spite of that, it allows an adjusted indicator of the flow of Italian tourists to be obtained in advance, with almost one month, as the model examples showed.

Considering that the tourism significance for Romanian economy is very important, future steps to be taken involve the creating of similar analysis for tourists from other countries, in order to be improved the predictive quality of the model.

With time and the continuing growth of Internet usage the results of this procedure can only enhance in the future.

Disclaimer

Responsibility for the information and views expressed therein lies entirely with the author.

References

- Baker, S., & Fradkin, A. (2011). *What drives job search? Evidence from Google search data. Technical report*, [Online] (URL http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1811247).
- Castle, L., Fawcett, W. P., & Hendry, F. (2009). *Nowcasting is not just contemporaneous forecasting*. National Institute Economic Review, pp. 71-89, URL <http://ner.sagepub.com/content/210/1/71.abstract>.
- Concha, A., & Galán, E. (2012). *Tracking the Future on the Web: Construction of Leading Indicators Using Internet Searches*, BANCO DE ESPAÑA, Madrid, 2012
- Choi, H., & Varian, H. (2009). *Predicting the present with Google Trends. Technical report*, Google. [Online] (URL http://google.com/googleblogs/pdfs/google_predicting_the_present.pdf).
- Choi, H., & Varian, H. (2011). *Predicting the present with Google Trends*. [Online] (URL <http://people.ischool.berkeley.edu/~hal/Papers/2011/ptp.pdf>).
- Ettredge, M., Gerdes, J., & Karuga, G. (2005). *Using web-based search data to predict macroeconomic statistics*. Communications of the ACM, pp. 87-92. URL <http://portal.acm.org/citation.cfm?id=1096010>.
- Holmes, E.E., Ward, E.J., & Scheuerell, M.D. (2014). *Analysis of multivariate time-series using the MARSS package*. [Online] (URL <http://cran.r-project.org/web/packages/MARSS/vignettes/UserGuide.pdf>).
- Lazer, D. M., Kennedy, R., King, G., & Vespignani, A. (2014). *The parable of google flu: traps in big data analysis*.
- McIver, D.J., & Brownstein, J.S. (2014). *Wikipedia Usage Estimates Prevalence of Influenza-Like Illness in the United States in Near Real-Time*. PLoS Comput Biol 10(4): e1003581
- McLaren, N. & Shanbhoge, R. (2011). *Using internet search data as economic indicators. Bank of England Quarterly Bulletin*. [Online] (URL <http://www.bankofengland.co.uk/publications/quarterlybulletin/qb110206.pdf>).
- Perduca, V., & Ferreira P. (2014). *Improving prediction of unemployment statistics with Google trends*.
- Online database <http://ec.europa.eu/eurostat/>.
- <http://ec.europa.eu/>.
- <http://www.wttc.org/eng/>.

