



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA

CARMA 2016

**1st International Conference on Advanced Research
Methods and Analytics**

July 6 – 7, 2016 · Valencia, Spain

Congress UPV

Proceedings of the 1st international Conference on Advanced Research Methods and Analytics, CARMA2016.

The contents of this publication have been evaluated by the Scientific Committee which it relates and the procedure set out <http://www.carmaconf.org/>

Scientific Editors

Josep Domenech
Alicia Mas-Tur
Norat Roig-Tierno
Maria Rosalia Vicente

© of the texts: authors

© 2016, Editorial Universitat Politècnica de València
www.lalibreria.upv.es / Ref.: 6287_01_01_01

ISBN: 978-84-9048-462-3 (print version)

Print on-demand

DOI: <http://dx.doi.org/10.4995/CARMA2016.2016.4355>



1st international Conference on Advanced Research Methods and Analytics, CARMA2016.

This book is licensed under a [Creative Commons Attribution-NonCommercial-NoDerivatives-4.0 International license](https://creativecommons.org/licenses/by-nc-nd/4.0/)

Editorial Universitat Politècnica de València <http://ocs.editorial.upv.es/index.php/CARMA/CARMA2016>

First International Conference on Advanced Research Methods and Analytics

Preface

This volume contains the selected papers of the First International Conference on Advanced Research Methods and Analytics (CARMA 2016), which was held in Valencia, Spain, during July 6th and 7th of 2016. Research methods in economics and business are evolving with the increasing availability of comprehensive sources of data. As these methods are becoming more interdisciplinary, CARMA 2016 provided researchers and practitioners with a forum to exchange ideas and advances on how emerging research methods are applied to different fields of social sciences as well as to discuss current and future challenges.

The scientific program was divided into three tracks. Track 1 focused on **Web and Big Data in Social Sciences** and included 23 papers dealing with topics such as Big Data in official statistics, Internet econometrics, geospatial and mobile phone data, and public opinion mining. All received papers were peer-reviewed by two or three members of the scientific committee, led by Dr. Josep Domenech and Dr. María Rosalía Vicente. The track also featured a special session on “Big data and nowcasting macroeconomic indicators” organized by Eurostat.

Track 2 and Track 3 dealt with **Qualitative and Comparative Methods** and **Advanced Regression Methods**. These sessions, comprising 10 papers, stand out on social innovation, technology transfer, entrepreneurial activities or business research. As with Track 1, two or three members of the scientific committee led by Dr. Alicia Mas-Tur and Dr. Norat Roig-Tierno performed peer-reviews of every manuscript.

CARMA 2016 also featured two keynote speakers that overviewed important and current topics. The opening keynote speech was delivered by Dr. Antonino Virgillito, head of the “Business Intelligence, Mobile and Big Data Architecture” unit at the Italian National Statistical Institute (Istat). His talk highlighted the challenges, experiences and future steps of Big Data in official statistics. The closing keynote speech was given by Prof. Sascha Kraus, Full Professor and Chairholder in Strategic Management and Entrepreneurship at the University of Liechtenstein. His talk dealt with the usage of fsQCA in innovation and entrepreneurship research.

The conference was hosted by the Faculty of Business Administration and Management of the Universitat Politècnica de València, which has been recently ranked as the best technical university in Spain by the Academic Ranking of World Universities (ARWU)

2015. Valencia is a city of culture and heritage. It is the third largest city in Spain and its location by the Mediterranean Sea provides their citizens and visitors with a privileged weather.

The organizing committee would like to thank all who made the first edition of CARMA a great success. Specifically, thanks are indebted to the invited speakers, authors, scientific committee members, reviewers, session chairs, presenters, sponsors, supporters and all the attendees. Our final words of gratitude must go to the Faculty of Business Administration and Management of the Universitat Politècnica de València for supporting CARMA 2016.

Josep Domenech

Alicia Mas-Tur

Norat Roig-Tierno

Maria Rosalia Vicente

Conference Committee

Conference Co-Chairs

Josep Domenech, Universitat Politècnica de València
Alicia Mas-Tur, Universitat de València
Norat Roig-Tierno, ESIC Business & Marketing School

Conference Secretary

Desamparados Blazquez, Universitat Politècnica de València

Sponsors

BigML
DevStat

Supporters

Universitat Politècnica de València
Facultad de Administración y Dirección de Empresas
Departamento de Economía y Ciencias Sociales

Scientific committee

Track 1: Web & Big Data in Social Sciences

Track chairs: Josep Domenech, Universitat Politècnica de València
María Rosalía Vicente, Universidad de Oviedo

Isidro Aguillo, Consejo Superior de Investigaciones Científicas (CSIC)
Concha Artola, Banco de España
Nikolaos Askitas, Institute for the Study of Labor (IZA)
Giulio Barcaroli, Italian National Institute of Statistics
Marc Bogdanowicz, Institute for Prospective Technological Studies (JRC-IPTS)
Pablo de Pedraza, Universidad de Salamanca / University of Amsterdam
Giuditta de Prato, Institute for Prospective Technological Studies (JRC-IPTS)
Benjamin Edelman, Harvard Business School
Daniel Gayo-Avello, Universidad de Oviedo
José A. Gil, Universitat Politècnica de València
Felix Krupar, Alexander von Humboldt Institute for Internet and Society
Rocío Martínez Torres, Universidad de Sevilla
Suzy Moat, University of Warwick
Esteban Moro, Universidad Autónoma de Madrid / Universidad Carlos III

Maurizio Naldi, Università di Roma Tor Vergata
Michaela Nardo, Joint Research Centre (JRC)
Bülent Özel, Universität Jaume I
Ana Pont, Universitat Politècnica de València
Tobias Preis, University of Warwick
Ravichandra Rao, Indian Statistical Institute
Fernando Reis, Eurostat
Andrea Scharnhorst, Royal Netherlands Academy of Arts and Sciences
Vincenzo Spiezia, OECD
Pål Sundsøy, Telenor
Sergio L. Toral Marín, Universidad de Sevilla
Michela Vecchi, Middlesex University

Track 2: Qualitative and Comparative Methods

Track chair: Norat Roig-Tierno, ESIC Business & Marketing School

Adrian Dusa, University of Bucharest
Ian Jenson, University of Tasmania
Kati Kasper-Brauer, Freiberg University of Technology
Alexander Leischnig, Bamberg University
Jordi Paniagua, Universidad Católica de Valencia
Domingo Riberio-Soriano, Universitat de València
Carsten Q. Schneider, Central European University

Track 3: Advanced Regression Methods

Track chair: Alicia Mas-Tur, Universitat de València

Hervé Abdi, The University of Texas at Dallas
José Antonio Belso, Universidad Miguel Hernández
Jasmina Berbegal, International University of Catalunya
Ricarda B. Bouncken, University of Bayreuth
Gabriel Cepeda, Universidad de Sevilla
José Manuel Guaita, Valencian International University
Alenka Janko Spreizer, University of Primorska
José M. Merigó, University of Chile
Marcin W. Staniewski, University of Finance and Management in Warsaw

Local Organization

Andrea Burgos Mascarell

Sergi Doménech de Soria

Alejandro Gil Ródenas

Andrea Rey-Martí

Web & Big Data in Social Sciences

Some guidance for the use of Big Data in macroeconomic nowcasting

Mazzi, Gian Luigi^{a1}

^aEurostat, European Commission, Luxembourg.

Abstract

This paper develops an operational step by step approach aiming to facilitate the use of Big Data in nowcasting exercises. Each step includes a description of the problem and a set of recommendations addressing the most relevant available solution. The approach includes nine steps starting from the theoretical availability of Big Data until the publication of new nowcasting including also Big Data. In designing this operational step by step approach, the preliminary results of an ongoing Eurostat project on Big Data and macroeconomic nowcasting have been used as a starting point. Further elaboration has been carried out in order to make the operational step by step approach more concrete and prescriptive. Its aim is to provide a concrete help for experts involved in the construction of nowcasting especially in the judgment about the usefulness of the presence of Big Data in their models. It also provides guidance related to the dissemination of new nowcasting based also on Big Data.

Keywords: Big Data; Nowcasting.

¹ The information and views set out in this paper are those of the author and do not necessarily reflect the official opinion of the European Commission.

1. Introduction

The availability of Big Data is opening new challenging ways of producing statistics. Big Data can be particularly relevant to increase the timeliness of macroeconomic indicators by means of new types of nowcasting. At present, Big Data should be viewed realistically more as a complement of traditional information to produce nowcasting instead of an alternative. The presence of Big Data can substantially change the traditional ways of building up nowcasting. This paper, also based on the preliminary results of an ongoing project on Big Data and macroeconomic nowcasting, is proposing a step by step approach for the utilization of Big Data in a nowcasting exercise.

2. A step by step approach for the use of Big Data in a nowcasting exercise

2.1. The step by step approach

In this section we are proposing some guidance, expressed in a form of a step by step approach, for using Big Data when building up macroeconomic nowcasting. Each step will be accompanied by a detailed description as well as one or more recommendations, suggested to the compilers of macroeconomic nowcasting. Table 1 summarises the various steps together with their aim while, in the following subsection, each step will be further detailed.

Table 1.

Steps	Title	Aim
Step 0	Big Data usefulness within a nowcasting exercise	Checking for the existence of adequate Big Data sources
Step 1	Big Data search	Identification of the appropriate Big Data
Step 2	Availability and quality	Verification of Big Data availability and its quality
Step 3	Accounting for Big Data specific features	Move from unstructured Big Data to a structured dataset
Step 4	Big Data pre-treatment	Removing deterministic and periodic undesirable effects
Step 5	Presence of bias	Checking Big Data for the presence of bias

Step 6	Big Data modelling	Identifying the best modelling strategy
Step 7	Results evaluation of Big Data based nowcasting	Checking for the effective contribution of Big Data
Step 8	Implementing Big Data based nowcasting	Timing and scheduling for the new nowcasting

2.2. Big Data usefulness within a nowcasting exercise

2.2.1. Description

This first step should investigate the potential usefulness of Big Data for a specific indicator of interest, such as GDP growth, inflation or unemployment rate. Big Data sources should be considered for their ability of improving the overall quality of existing nowcasting or of producing timelier estimates. The theoretical soundness of the relationships between existing Big Data sources and the target variables should also be investigated.

2.2.2. Recommendations

- Suggest the use of Big Data only when there are well founded expectations of their usefulness either for fixing problems in existing nowcasting or to improve the timeliness.
- Do not consider Big Data sources with doubtful or even spurious correlations with the target variable.

2.3. Big Data search

2.3.1. Description

Once Big Data passes the “need check” in the previous step, the next action of the Big Data based nowcasting exercise is a careful search for the specific Big Data to be collected. There are many potential providers such as social networks, traditional business systems, the Internet of things, etc. It is very difficult to give general guidelines on a preferred data source because the choice is heavily dependent on the target indicator of the nowcasting exercise.

2.3.2. Recommendations

- Searching in the wider possible set of Big Data having clearly in mind the specificities and the characteristics of the target variable as well as what we want to nowcast.
- Checking for the adherence of available Big Data to what the target variable is really measuring.

2.4. Availability and quality

2.4.1. Description

Having identified the preferred source of Big Data, the second step requires assessing the availability and quality of the data. A relevant issue is whether direct data collection is needed, which can be very costly, or a provider makes the data available. In case a provider is available, its reliability and cost should be assessed, together with the availability of meta data, the likelihood that continuity of data provision is guaranteed, and the possibility of customization (e.g., make the data available at higher frequency, with a particular disaggregation, for a longer sample, etc.). All these aspects are particularly relevant in the context of applications in official statistical offices. As the specific goal is nowcasting, it should be also carefully checked that the temporal dimension of the Big Data is long and homogeneous enough to allow for proper model estimation and evaluation of the resulting nowcasts.

2.4.2. Recommendations

- Privileging data providers which are able to give sufficient guarantee of the continuity of the data process and of the availability of a good and regularly updated metadata associated to the Big Data
- Privileging Big Data sources which ensure sufficient time coverage to properly building up a nowcasting exercise.

2.5. Accounting for Big Data specific features

2.5.1. Description

The third step analyzes specific features of the collected Big Data. A first issue concerns the amount of the required storage space and the associated need of specific hardware and software for storing and handling the Big Data. The second issue is the type of the Big Data, as it is often unstructured and may require a transformation into cross-sectional or time series observations.

2.5.2. Recommendations

- Creating a Big Data specific IT environment where the original data are collected and stored with associated routines to automatically convert them into structured, either cross-sectional or time-series datasets.
- Ensure the availability of an exhaustive documentation of the Big Data conversion process.

2.6. Big Data pre-treatment

2.6.1. Description

Even when already available in numerical format or after their transformation into numerical form as in the previous step, pre-treatment of the Big Data is often needed to remove deterministic patterns such as outliers and calendar effects and deal with data irregularities, like missing observations. Furthermore, seasonal and non-seasonal short-term movements (i.e. infra-monthly ones) should be removed accordingly to the characteristic of the target variable. Since not all seasonal and calendar adjustment methods can be applied when data are available at high frequency, appropriate adjustment techniques need to be identified when the data are available at high frequency. The size of the datasets suggests resorting to robust and computationally simple univariate approaches.

2.6.2. Recommendations

- Whenever possible, all the data treatment described in this step should be done within a unique framework in order to avoid inconsistencies between different parts of the process.
- The filtering of Big Data should be consistent to the one used for the target variables: for example if the target variable is not seasonally adjusted, there is no reason to remove the seasonal component from Big Data and vice-versa.

2.7. Presence of bias

2.7.1. Description

This step requires assessing the presence of a possible bias in the answers provided by the Big Data, due to the so-called “digital divide” or the tendency of individuals and businesses not to report truthfully their experiences, assessments and opinions. Another relevant and partially related problem, particularly relevant for nowcasting, is the possible instability of the relationship with the target variable. This is a common problem also with standard indicators and traditional nowcasting exercises. Both issues can be however tackled at the modelling and evaluation stages.

2.7.2. Recommendations

- If a bias in the Big Data answers is observed, provided that it has been reasonably stable in the last few years, a bias correction can be included in the nowcasting strategy.
- If a bias in the Big Data answers is very unstable, then the Big Data should be considered not reliable enough to be used in a nowcasting exercise.
- In order to deal with a possible instability of the relationships between the Big Data and the target variables, nowcasting models should be re-specified on a regular basis (e.g. yearly) and occasionally in presence of unexpected events.

2.8. Big Data modelling

2.8.1. Description

This step requires the identification of the most appropriate econometric technique when building up a nowcasting exercise with Big Data. It is important to be systematic about the correspondence between the nature and the size of the selected Big Data and the method that is used. There are a number of dimensions along which we wish to differentiate.

In the first one we address the choice between the use of methods suited for large but not huge datasets, and therefore applied to summaries of the Big Data (such as Google Trends, commonly used in nowcasting applications), or of techniques specifically designed for Big Data. For example, nowcasting with large datasets can be based on factor models, large BVARs, or shrinkage regressions.

Huge datasets can be handled by sparse principal components, linear models combined with heuristic optimization, or a variety of machine learning methods, such as LASSO and LARS regression which, though, are generally developed assuming i.i.d. variables. It is difficult to provide an a priori ranking of all these techniques and there are few empirical comparisons and even fewer in a nowcasting context, so that it may be appropriate to apply and compare a few of them for nowcasting the specific indicator of interest. In absence of a multifrequency problem, those techniques can work for variable selection or data reduction as well as for the estimation of the nowcasting of the target variable.

In the second dimension we address the problem of the frequency of the available data. If this frequency is mixed, then specific techniques for mixed frequency data become relevant after having selected the variables or having reduced the dimension of the variables space accordingly with the techniques discussed above. Among mixed frequency models, UMIDAS stands out but also Bridge models can deserve a certain attention. UMIDAS provides a very flexible framework of analysis and can be adapted to work together with most if not all Big Data methods be they machine learning or econometric.

2.8.2. Recommendations

- In absence of any a priori information on the relative performance of various techniques, as many methods as possible should be evaluated and compared in a nowcasting context in order to select the best performing one.
- Alternative modelling strategies should be compared also by looking at the balance between their complexity in computational terms and their empirical performance.
- In case of mixed frequency data, linear methods such as UMIDAS and, as a second best, Bridge, should be privileged.
- Forecast combination and model averaging techniques, also when the mixed frequency aspect is present, can be used as an alternative to a large-scale comparison among competing techniques.

2.9. Results evaluation of Big Data based nowcasting

2.9.1. Description

The final step consists of a critical and comprehensive assessment of the contribution of Big Data for nowcasting the indicator of interest. This should be carried out within a real-time or a pseudo-real time exercise. In order to avoid, or at least reduce the extent of, data and model snooping, a cross-validation approach should be followed, whereby various models and indicators, with and without Big Data, are estimated over a first sample and they are selected and/or pooled according to their performance, but then the performance of the preferred approaches is re-evaluated over a second sample.

This procedure provides a reliable assessment of the gains in terms of enhanced nowcasting performance from the use of Big Data. For some critics about the usefulness of Big Data see Hartford (2014) and Lazer et al. (2014).

2.9.2. Recommendations

- Conducting an in-depth real-time or pseudo real-time simulation of competing models in order to evaluate their relative performance in nowcasting the variable of interest.
- Models including Big Data should be preferred when they significantly lead to an improvement of the reliability and accuracy of the nowcasting at the same point in time.
- Models including Big Data should also be preferred when they allow for timelier nowcasting without any significant loss in terms of reliability and accuracy.

2.10. Implementing Big Data based nowcasting

2.10.1. Description

In case the in-depth comparative analysis carried out in the previous steps suggests that the use of Big Data can improve the nowcasting for a given variable of interest, they can be then implemented. At this stage, the institution in charge of producing nowcasting should take several relevant decisions related to the number of the nowcasting to be implemented and their scheduling. For example, it is possible to decide to publish just one nowcasting (e.g. at the very end of the reference period or at the very beginning of the following one), to produce two nowcastings (e.g. one in the middle of the reference period and one at the very end), or to produce a sequence of nowcasts scheduled at weekly or even daily frequency. Such decisions should take into account, among other, the trade-off between timeliness and reliability, the user needs as well as some more institutional considerations.

2.10.2. Recommendations

- Implementing and publishing the most reliable nowcasts available either at the end of the reference period or at the beginning of the following one.
- Moving towards a daily or weekly update on nowcasting already during the reference period, only after detailed pros and cons analysis and a consultation of the most relevant stakeholders.
- The new Big Data based nowcasting should be accompanied by clear metadata and widely available reference and methodological papers.

3. Conclusions

In this paper, we have proposed a new operational step by step approach for using Big Data in a nowcasting exercise. It aims to facilitate the activity of experts involved in the construction of nowcasting by providing a set of recommendations associated to various operational steps.

References

- Hartford, T. (2014, April). Big data: Are we making a big mistake? Financial Times. Retrieved from <http://www.ft.com/cms/s/2/21a6e7d8-b479-11e3-a09a-00144feabdc0.html#ixzz2xcdlP1zZ>.
- Lazer, D., Kennedy, R., King, G. & Vespignani, A. (2014). The Parable of Google Flu: Traps in Big Data Analysis. *Science*, 143, 1203-1205.

Nowcasting with Google Trends, the more is not always the better

Combes, Stéphanie^a and Bortoli, Clément^b

^aDepartment of statistical methods, INSEE, France, ^bDepartment of conjuncture, INSEE, France.

Abstract

National accounts and macroeconomic indicators are usually published with a consequent delay. However, for decision makers, it is crucial to have the most up-to-date information about the current national economic situation. This motivates the recourse to statistical modeling to “predict the present”, which is referred to as “nowcasting”. Mostly, models incorporate variables from qualitative business tendency surveys available within a month, but forecasters have been looking for alternative sources of data over the last few years. Among them, searches carried out by users on research engines on the Internet – especially Google Trends – have been considered in several economic studies. Most of these exhibit an improvement of the forecasts when including one Google Trends series in an autoregressive model. But one may expect that the quantity and diversity of searches convey far more useful and hidden information. To test this hypothesis, we confronted different modeling techniques, traditionally used in the context of many variables compared to the number of observations, to forecast two French macroeconomic variables. Despite the automatic selection of many Google Trends, it appears that forecasts’ accuracy is not significantly improved with these approaches.

Keywords: *nowcasting; Google Trends; macroeconomics; high dimension; machine learning; time series.*

1. Introduction

Official statistics are often published within irreducible delays¹. But for decision makers, it is crucial to have access to the most up-to-date information about the current national economic situation. This is why developing some efficient forecasting tools and identifying the most relevant data sources are a serious issue in macroeconomics. Real time forecasting (or *nowcasting*) of macroeconomic indicators usually implies to incorporate variables from qualitative business surveys or sometimes financial variables. Over the last few years, forecasters have also been looking into data from the Internet and at trending searches made by Google users in particular.

In 2006, Google launched *Google Trends*, a tool that provides data series free of charge which reflect the interest of Internet users in a query or a set of semantically linked search terms. If this application has been popularized by advertising the most popular searches of the moment, it has also become a well-known source of data for economic studies. Characterized by their high frequency compared to official indicators and their short delay of publication (a week), they have been investigated by numerous economists over the last few years. Indeed, the global evolution of queries made by users about particular products or subjects via the search engine is likely to reflect the potential volume of sales of these products or the predominance of the subject for individuals at the time. These data could therefore be considered as indicators of consumer purchase intention or concerns (for example queries about unemployment benefit may give a hint of the evolution of the unemployment rate). Plus, the soaring penetration rate of equipment of households in computers and Internet connection makes them a credible source of information on individuals (less likely on companies).

The most famous use case of prediction with Google Trends is the Google Flu application developed by Google to forecast the spread of the flu epidemic in real time, based on user queries, in 2008. First launched in the United States, the tool was extended the following year to a dozen European countries, including France. In 2009, the group published an analysis of the benefits of using these series to forecast socio-economic indicators (Choi and Varian, 2009). According to this study, which used American data, forecasting automobile purchases, retail sales and purchases of dwellings could be improved by introducing this type of series into simple models using the dynamics of the series of interest (autoregressive model).

¹ In France, the main quantitative data available on household expenditure for example is the monthly household consumption expenditure on goods, published within one month and its equivalent in services is published within two months. Finally, an initial estimate of quarterly spending on all goods and services is published in the middle of the following quarter.

Askitas and Zimmermann (2009) used the frequency of use of certain search terms to forecast the unemployment rate in Germany; Kulkarni et al. (2009) suggested a link between the frequency of several search terms and housing prices in the United States; Vosen and Schmidt (2013) also used this type of series to forecast household expenditure in the United States.

Using Google Trends data in a variety of fields and incorporating them into more complex econometric models were also tested subsequently. It is along those lines that our study contemplates to contribute. Indeed Google Trends supplies a large pool of series that may convey useful yet hidden information. Automatic variables selection or extraction methods seem to match perfectly this situation where a lot of potential regressors are available, the number of observations is limited, and the expert may not want to constrain the specification of the model too much. In this study, we confronted several approaches used for forecasting in high dimension: variables selection techniques well-known in macroeconomics, variables extraction methods which aim at summarizing a large set of data in a smaller one, averaging methods to take into account the modeling uncertainty, and, eventually, non-parametric methods borrowed from machine learning, which appear to provide accurate predictions in numerous and various fields.

The next section describes more precisely our data and the treatments that were operated on them. Then, we remind our reader quickly with the concepts behind the different techniques we used, and eventually, we present and discuss the main results.

2. Data

The main attraction of the Google Trends data for the economic outlook lies in the fact that they can be mobilized quickly and at a higher frequency than most traditional economic series. Indeed, data related to one given week are published at the end of the very week. Data can also be filtered by geographic origin: we could therefore restrict our study to searches carried out in France. Available data are pretreated which means that raw series corresponding to the real frequency of use of a search term are not made public. Applied treatments are not very well documented but series are supposedly corrected accordingly to a trend resulting from an increase in popularity of the search engine itself. They are normalized too so that their maximum always equals 100, which means that they might be revised between one extraction at a certain date and another one later on and that direct comparison between two distinct series is not possible.

Google provides categories grouping queries by topics. More of one hundred of them are available organized in a three levels hierarchy. Normalization of categories differs from keyword's one: the frequency of the category in the first week of 2004 is used as a reference, the following points in the series are expressed as deviations from this level. Since the

meaning of a search term can evolve over time, it seems preferable to work on categories or concepts rather than on specific terms. Plus the strategy of choice of keywords would be very subject to subjectivity in addition to consequent manual task. For example, the "Sports" category aggregates all search terms linked with the field of sport. French Google users have shown an increased interest in this topic in the summers of even years (figure 1). Indeed searches related to sport showed a marked increase during the football World Cup 2006, 2010, 2014, the European football championships and the Olympic Games in the summers of 2004, 2008 and 2012. Purchases of televisions usually increase significantly at times of major sports events, so using the "Sports" category seems to be a natural choice to measure the degree of interest that a sports event can generate among French consumers.

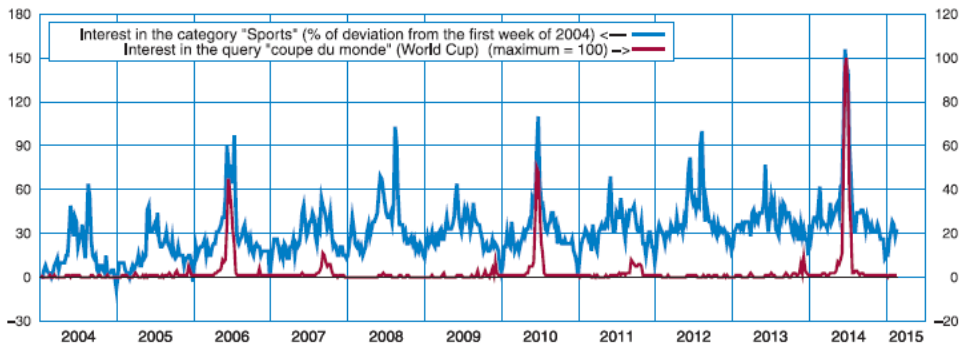


Figure 1. Examples of Google Trends chronicles for two chosen keywords. Source: Google Trends (2015)

In the context of our study, we selected a pool of 50 categories which may be correlated with the macroeconomic situation in one way or the other. The Google Trends categories were first transformed into a monthly format, weeks overlapping a month were distributed accordingly to the number of days in each month². Series were then seasonally adjusted; their monthly growth rates were computed to produce the explanatory variables, as well as their first time lag (i.e. the value of this growth rate in the previous month).

For this study, two targets have been considered: the household consumption in goods and the manufacturing production index. Representing more than half of GDP, household consumption is the largest item in final domestic demand, its estimate gives therefore a good outline of the whole activity. The first available data is the monthly household consumption expenditure on goods, published within one month. The publication of the manufacturing production index takes more time (two months), but its variation explains most of the quarterly GDP's evolution, it is then crucial to be able to produce accurate advanced estimates. In order to forecast these indices in real time or before they are published, usual

² This aspect makes it difficult to use techniques mixing data with different frequencies such as MIDAS (Mixed-data sampling), the advantage of the higher frequency was then not exploited here.

Para seguir leyendo haga click aquí