Contribution from Google Trends for forecasting the short-term economic outlook in France: limited avenues

Clément Bortoli

Département de la Conjoncture

Stéphanie Combes

Département des méthodes statistiques

Representing more than half of GDP, household consumption is the largest item in final domestic demand. It is therefore essential to have information that is as up to date as possible on any changes in order to define and forecast activity in France. The main quantitative data available on household expenditure is the monthly household consumption expenditure on goods, published within just one month. Turnover indices, giving information in particular on spending on services, are available within two months. Finally, an initial estimate of quarterly spending on all goods and services is published in the middle of the following quarter.

In order to estimate spending in real time or before these figures are published, the usual forecasting models often incorporate variables from qualitative business tendency surveys. These data are available within the month, and make up the early indicators for a certain number of macroeconomic variables.

Over the last few years, forecasters have also been looking at data from the Internet and in particular at trending searches by Google users as an information source that could perhaps improve their forecasts. As well as being free, the speed at which these data can be mobilised makes them a very attractive proposition for studying economic outlook: they are available at the end of each week, and show the popularity of Internet users' queries virtually in real time. In addition, the volume of queries made by users about particular products via the search engine could reflect the potential volume of sales of these products. These data could therefore be considered as indicators of consumer purchase intention, both for manufactured goods and for services.

Google provides details of trending searches being made by users through their search tool Google Trends. The categories offered by this tool group queries by topic. There are a great many that could prove relevant for forecasting household consumption; however, if they were to be used in the context of econometric models we would have to be in a position to judge the quality of their contribution to the forecast. In this paper we look at the methods of combining models which would enable us to take advantage of all available information and to take into account forecasts obtained from different models with similar performances, using a «Bayesian» approach.

From the different models tested, it can be seen that using Google trending searches improves the forecasting of household expenditure in only a limited way. Specifically, taking these trends into account would not improve the overall forecast of monthly household consumption expenditure on goods or services, as their level of heterogeneity is very high. On the other hand, results obtained for purchases of certain goods (especially clothing and household durables) are more positive, and some Google Trends categories can improve forecasting for these items. However, even in these cases, any improvement is fairly limited.

Changes in household spending set the compass for the French economy

Household consumption represents the largest item in final domestic demand in France

In France, final household consumption of goods and services is the largest item in final domestic demand, representing about half in current euros since the beginning of the 1980s, with the other half being spread over other types of expenditure (individual and collective consumption by general government and non-profit institutions serving households), and gross fixed capital formation, which represent one quarter each. The share of household consumption in value in GDP has been relatively stable since the beginning of the 1980s, varying between 52% and 56%.

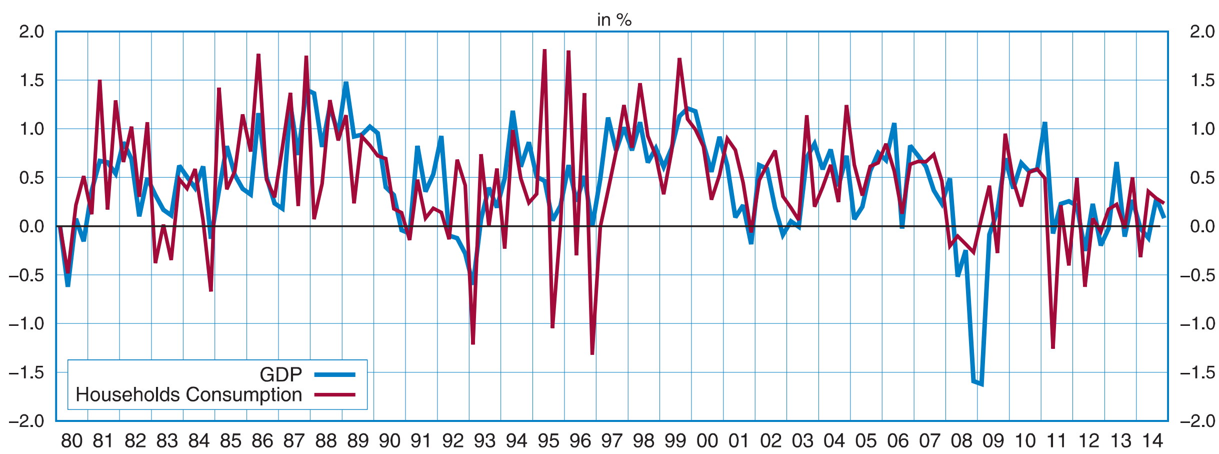
Household consumption expenditure plays a key role in short-term changes in the economy

The different macroeconomic equations used to model consumption behaviour take many factors into account (*Faure et al., 2012*). Firstly, household consumption reacts to fluctuations in purchasing power. There are other factors that can influence spending, such as unemployment, due to behaviours relating to"precautionary savings", price fluctuations or interest rates. Secondly, one-off events can cause jolts in consumer expenditure. For example, when temperatures differ significantly from seasonal norms in autumn or spring, this can lead to major disparities in spending on heating (e.g. contributing -0.2 points to the increase in total consumer expenditure in Q4 2014, and -0.1 points to GDP growth). Similarly, car-scrapping bonuses and successive adjustments to the ecological "bonus-malus" system introduced in 2008 caused major jolts in automobile consumption (contributing -0.8 points to total growth in consumption expenditure in Q2 2011). All in all, household consumption often has a somewhat uneven profile. Because of its influence on economic activity in France, it is one of the main contributory factors to variance in quarterly growth ("volatility") of GDP (*Graph 1*). Between 1980 and 2014, it accounted for 37% of this variation, on average, which was less than investment (40%), but more than inventory change (22%).

Household consumption of goods accounts for a very large proportion of the volatility of total expenditure

The volatility of total household consumption is mainly the result of consumption of goods, which accounts on average for 84% of this total volatility since 1980. The structure of the consumption of goods has changed little in thirty years. It is spread over food products (36% of consumption of goods on average), energy consumption (17% on average, which covers both spending on heating and purchase of fuel) and purchase of manufactured goods (47% on average since 1980), especially automobiles, household durables (computers and electronic

1 – GDP growth rate and household consumption



Source: INSEE

products, domestic electrical appliances and furniture) and clothing. Automobile purchases evolve in a particularly uneven fashion mainly because of car-scrapping bonuses and ecological "bonus-malus" schemes (see above); over the period under consideration, they accounted for 49% of the volatility of consumption of goods (*Table 1*), or 42% of the volatility of total household consumption.

The share of services in household expenditure has increased continuously since the beginning of the 1980s

Since the beginning of the 1980s, the share of household consumption given over to services has continued to grow. In 1980, household consumption expenditure on services represented only 40% of total expenditure (in current euros), whereas it now stands at 53%. This dynamic growth can also be seen in constant euros, though a little less marked: over the same period, the proportion of services increased from 46% to 54% in volume. A major proportion of the volatility of consumption of services is due to a relatively small number of items (*Table 2*): thus, expenditure on accomodation and catering, transport and information-communication services (which include mainly spending on telephones and Internet subscriptions) accounts for 56% of the volatility of household consumption of services, although it represents less than 30% of the average total since 1980. Conversely, real and imputed rents represent 33% of services in terms of value, but because of their inertia, they account for only 8% of the volatility of expenditure on services - or less than 1% of the volatility of total household spending.

Table1
Expenditure by type of product: weight in the consumption of goods and contribution to the variance in changes in the consumption of goods

	Weight in the consumption of goods	Contribution to the variance in changes in the consumption of goods
Food products	36%	11%
Engineered goods	47%	68%
including: Automobiles	*12*	*49%*
Household durables	*8%*	*7%*
Clothing	*12%*	*8%*
Energy	17%	21%
Total goods	100%	100%

How to read the table: between 1980 and 2014, expenditure on food products represented on average 36% of total household consumption of goods (in current euros). Over the same period, growth in this expenditure accounted for 11% of variance in quarterly changes in total household consumption of goods

Source: INSEE

Table 2
Expenditure by type of product: weight in the consumption of services and contribution to the variance in changes in the consumption of services

	Weight in the consumption of services	Contribution to the variance in changes in the consumption of
Housing services	33%	8%
Accomodation and catering	13%	26%
Information and communication	10%	17%
Transport services	6%	13%
Other services	38%	36%
Total services	100%	100%

Source: INSEE

Input from Google Trends for forecasting the short-term economic outlook in France: limited avenues

Online searches may be able to provide information quickly on household consumption

Consumer indices are gradually becoming available

Final household consumption expenditure on goods and services clearly plays a key role in short-term developments in the French economy. However, data on household spending are not instantly available. Only a few specific indicators are available in virtually real time: vehicle registrations can be provided from the first working day of the following month; data on electricity consumption are published daily, virtually in real time. The first aggregated data available for household expenditure is the monthly household consumption expenditure on goods, published about 30 days after the month in question. Turnover in services, providing information on expenditure on services in particular, are published almost 60 days after the month under consideration. Other data on services are provided later still, for example telephone data published by ARCEP about 70 days after the end of the quarter in question. Lastly, around 45 days after the end of a quarter, an initial estimate of the quarterly household expenditure accounts is produced (national accountants must extrapolate certain indicators that are not yet known at this date).

Online search statistics may be able to provide early information on consumption

Making an accurate and faster diagnosis of short-term changes in household consumption expenditure therefore requires having early indicators available, virtually in real time. With this in mind, using statistics based on online searches would seem to be a promising source, if only because the Internet plays an increasing part in household purchases. Indeed, between 2006 and 2011, the share of purchases of durable goods on the Internet rose from 2% to 9%, and was as high as 11% for cultural items (*Krankadler, 2014*). The Internet also plays an increasing part in the consumption of services (transport, business, financial services, etc.). Even if a purchase does not take place online, households can obtain information beforehand through search engines. These queries may therefore contain some revealing purchase intentions.

Internet has already been used to forecast socio-economic indicators for several years

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 search terms that are semantically linked. Notably, this application gives wide coverage to the most popular searches by users of its search engine, practically in real time. 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 data series into simple models using the dynamics of the series to be forecast (autoregressive model). Using Google Trends data in a variety of fields and incorporating them into more complex econometric models were also tested subsequently.

The most famous example of using Google Trends was the Google Flu application developed by Google to forecast the spread of the flu epidemic in real time, based on user queries. In economics, *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 (2011)* also used this type of series to forecast household expenditure in the United States.

Google Trends data can be mobilised quickly

The main attraction of the Google Trends data for the economic outlook lies in the fact that they can be mobilised quickly and at a higher frequency than most traditional economic series, as data are published at the end of every week. Data can also be obtained by geographic origin: we can therefore restrict our study to searches carried out in France to try and produce a diagnosis on changes in the economic outlook in France. Raw series that correspond to the real frequency of use of a search term are not made public. In fact, the data that are made available are corrected according to a trend probably resulting from an increase in popularity of the search engine itself, and are presented in the form of time series of whole numbers, which are ultimately normalised so that their maximum equals 100 (*Graph 2*).

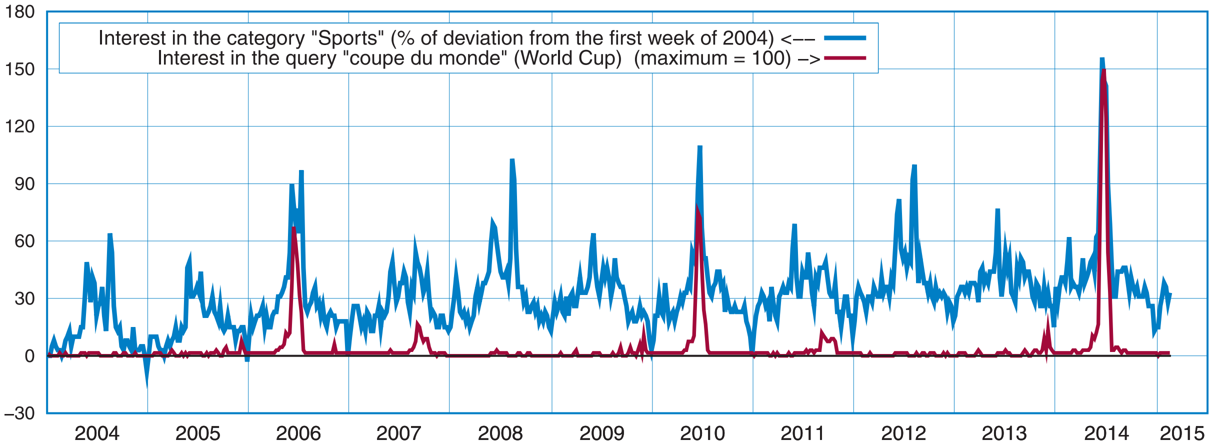
Google Trends categories aggregate similar terms

Since the meaning of a search term can evolve over time, it is preferable to work on categories or concepts rather than on specific terms. Google Trends provides users with statistics on sets of terms that are semantically close, called "categories". The normalisation that Google applies to the series relating to the Google Trends categories is different from that applied to simple queries: the frequency of the category in the first week of 2004 is used as a reference, with the following points in the series being expressed as deviations from this level. For example, the "Sports" category aggregates all search terms linked with the field of sport. French Google users have shown a heightened interest in this topic in the summers of even years. To be more precise, 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 sporting events, so using the "Sports" category can serve to measure the degree of interest that a sports event can generate among French consumers and hence to quantify this extra consumption of household equipment goods in real time.

The Google Trends tool nevertheless has several weaknesses

However, the way in which Google produces the data series disseminated on Google Trends lacks transparency, which is one of the weaknesses of this tool. The processing and the normalisations that are applied to queries and categories are not specified. Also, the management of categories and their composition over time, especially when a popular new query appears at a given date, are not documented. In addition, the tool has potentially damaging limitations in the production of statistics with long-lasting usability (*Box 1*). First, the series provided are the result of random sampling and can therefore differ from one data extraction to another. Next, the Google Trends application is, by its construction, dependent on developments in the marketing strategy of the group and on technological developments in the search engine itself, as adapted to meet the needs of its users. Since it was first created, the tool and the range of series to which the user has access have therefore changed substantially. Finally, the search engine has evolved a great deal: these changes are probably the reason for a marked deterioration in the performance of the Google Flu tool from 2009.

2 - Interest in the query "coupe du monde" (World Cup) and in the "Sports" category in Google Trends in France



Source: Google Trends

Box 1 - The long-lasting validity of series obtained with Google Trends is uncertain

Producing statistics on a regular basis requires that the data used are sustainable and of good quality. The data in Google Trends must be analysed from this standpoint. In particular, the series that are provided derive from random sampling and may therefore differ from one extraction to another. Specifically, with each data extraction, the tool provides a series constructed from the breakdown of queries from the history of a sample of users. Different samples may therefore produce non-coincident series for the same search request. This phenomenon has a tendency to increase when the query is an unusual one; however, the query frequency that is sufficient for the series to be relatively robust under the sampling technique is not provided. We can hypothesise that using the categories suggested by Google Trends, which group together a large number of queries under the same heading, will minimise noise in the data.

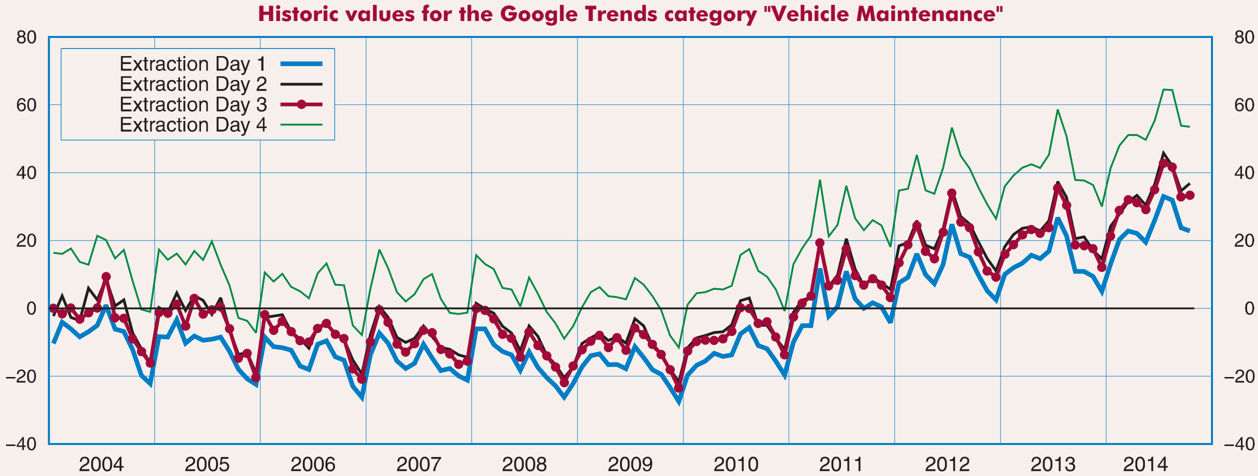
The question as to the future of the tool itself must also be asked. Sustainable production of an indicator from Google trending searches is likely to be under threat not only from changes in the group's business strategy but also from technological changes to the search engine. The Google Flu tool, another Google application, which at first proved to be effective in forecasting the spread of the flu epidemic, later came up against limitations which were attributed in part to these two types of change.

Google Flu was launched in 2008 in the United States and was extended the following year to a dozen European countries, including France. To produce the indicator, Google sought to identify from among thousands of keywords those that were most correlated with changes in official indicators provided by health monitoring bodies, such as statistics produced by clinical surveillance systems, which in the case of France was the national institute for health and medical research (INSERM). Search terms that had use-frequency peaks which were identical to the progress of the seasonal flu were selected. When this tool was launched in the United States, it looked very promising, since it provided an indicator that was one to two weeks ahead of the official publications, taking advantage of the almost immediate availability of the Google queries (*Ginsberg et al., 2009*). However, the tool was proved wrong for the first time in 2009 when it was unable to predict the non-seasonal epidemic of swine flu (H1N1) in the United States. Although it was corrected after this episode, Google Flu also experienced failures during the winters of 2011-2012 and 2012-2013, by widely overestimating the flu epidemics in the United States. This loss in performance was analysed by *Lazer et al. (2014)*, who put forward several explanations:

- the search behaviour of Internet users may be modified by media coverage of the epidemics being considered;
- the performance of the search engine and the algorithms may evolve and lead to a change in the way in which users use it.

Thus the performances of an indicator based on this tool are closely linked with Internet use habits. For example, the growing share of smartphone applications in Internet use could eventually lead to a reduction in the part played by search engines if Internet users favour applications dedicated to purchasing: the ability of trending searches to capture consumer behaviour may then be reduced. ■

Historic values for the Google Trends category "Vehicle Maintenance"



How to read it: Google Trends weekly data are shown in monthly forma.

Source: Google Trends

Input from Google Trends for forecasting the short-term economic outlook in France: limited avenues

The Google Trends categories broadly cover the different items of household expenditure

Among the categories offered by Google Trends, there are about fifty that could be potentially interesting candidates for forecasting the consumption of goods or services. Furthermore, these categories seem to cover all areas of expenditure (*Tables 3 and 4*).

By combining models, the diversity of series that can be mobilised for forecasting can be taken into account

The high number of potential explanatory variables requires a selection strategy ...

As well as fast data availability, which could mean that real economic activity can be evaluated with a much shorter delay than with most economic indicators, Google Trends has many data series and therefore must represent a major repository of information. Models could even incorporate time-lagged series by hypothesising that users are making queries up to one month before making their purchase. About fifty Google Trends series were targeted, producing about a hundred potential explanatory variables, for around 130 monthly observations, from January 2004 until the present. The number of potential regressors is therefore high compared with the number of observations. In practice, when forecasting, it is preferable to focus on the most parsimonious models, i.e. those which use only a limited number of variables. There are several strategies available for selecting those which are most relevant for the model.

Table 3
Examples of categories relevant in principle for forecasting household consumption of goods

	Manufactured goods				
		Durable goods			
Food	**Others manufactured goods**	Other durable goods	**Automobiles**	**Household durables**	Clothing
Tobacco products	Games		Autos and vehicles	Computers and electronics	Apparel
Alcoholic beverages	Toys		Vehicle shopping	Internet and Telecom	Sporting goods
Non Alcoholic beverages	Health		Vehicle maintenance	Consumer electronics	
Grocery and food retailers	Shopping		Vehicle brands	Mobile and wireless	
Mass merchants and Department stores	Beauty and fitness		Vehicle parts and accessories	Laptops and notebooks	
	Sports		Motorcycles	Home appliances	
	Sporting goods		Scooters and mopeds	Home furnishings	
	Kitchen and dining			Home and garden	
	Books and literature			Sports	
	Makeup and cosmetics				
	Luggage and travel accessories				

Table 4
Examples of categories relevant in principle for forecasting household consumption of services

Transport	Accommodations and food services	**Information et communication**	**Financial services**	Real estate	**Household services**
Travel	Restaurants	Internet and Telecom	Insurance	Apartment and residential rentals	Arts and entertainment
Car rental and taxi services	Hotels and Accommodations	Service providers	Banking		Hobbies and leisure
Air travel		Mobile and wireless			Beauty and fitness
Bus and rail		Books and Literature			Sports
		Shopping portals and search engines			

Input from Google Trends for forecasting the short-term economic outlook in France: limited avenues

... which in this case, cannot rely on expert judgment

The first involves selecting beforehand the most relevant variables for the study, based on expert judgement. For series that correspond to Google Trends categories, this expertise has yet to be constructed. Some categories may seem naturally to be of interest (e.g. "Travel", "Air travel" or "Bus and Rail" for spending on transport services), or one might prefer not to express any preconceptions on the selection of all the relevant variables.

Automatic selection of these series may be put in placebased on statistical criteria

Many econometric or data-analysis approaches have been developed to produce an automatic selection of explanatory variables from statistical criteria:

- iterative algorithms which add (respectively remove) a certain number of variables from an initial empty (respectively full) model, on the basis of a significance criterion,
- approaches to minimise an objective function with a penalty that favours parsimony (e.g. LASSO[1] or information criteria),
- approaches using principal component analysis, which can summarise information from a large number of variables into a small number of factors,
- combinations of models.

Combining models improves robustness

Combining models (*Box 2*) has the advantage of taking into account any uncertainty associated with the selected model, especially when several models have forecasting performances that are considered similar and it would be arbitrary to favour only one. In addition, these models provide forecasts which may differ significantly for the period being considered. Combining the forecasts produced by these different models better takes account of all available information, but is also more robust in the event of a shock produced by an isolated variable.

The relative weights of the forecasts produced by the models selected for the combination can be calculated in various ways, here for example using the Bayesian approach, which allows us to steer the model size towards parsimony.

Using Google trending searches, forecasts of household consumption can be improved slightly for certain products

Using Google Trends can improve consumption forecasts for certain products slightly

Using Google trending searches did not improve the forecasting of monthly household consumption expenditure on goods or services when this spending was aggregated, probably because the items are so very varied. On the other hand, results for the purchase of certain goods (clothing, household durables, food) and some services (transport) were more positive.

The Google Trends categories were used here to predict growth in monthly household consumption expenditure on goods (by volume) published by INSEE at the end of the month following the month under consideration. The Google Trends categories were transformed into a monthly format, then deseasonalised; their monthly growth rates were then used as potential explanatory variables, as well as their first time lag (i.e. the value of this growth rate in the previous month). In addition, the first four time lags of the dependant variable were also selected. A similar procedure was also applied to try to predict monthly growth in consumer expenditure on services. Estimates were made for the period March 2004 - December 2011 then the quality of the forecast was measured for the period January 2012 - December 2014 using the Root Mean Square Error (RMSE) criterion.

(1) The LASSO method corresponds to a linear regression which considers both the quality of the adjustment and the absolute value of the coefficients. The relative importance of these two objectives is steered by one parameter: if this is high, then many coefficients will be zero and thus many explanatory variables will not fit into the regression.

Input from Google Trends for forecasting the short-term economic outlook in France: limited avenues

Using Google trending searches did not improve forecasts for consumption expenditure on goods or services when this spending was considered at an aggregated level...

For total consumption of goods, the quality of results when forecasting with Google Trends data was no better than with a simple model using only the dynamics of the series (ARMA), whatever approach was used. Indeed, the most relevant variable to account for monthly growth in consumption of goods proved to be the growth of this same consumption in the previous month. When applying a Bayesian approach, Google Trends categories did indeed appear among the explanatory variables that frequently came up, but they were extremely varied ("Vehicle maintenance", "Consumer electronics", "Home furnishings", "Health", etc.). The difficulty in using these categories to explain in detail any changes in consumption of goods at an aggregated level lay partly in the highly varied nature of this consumption: generally speaking, the dynamics of spending on automobiles is going to be very different from spending on food products. Similar results were obtained for monthly consumption of services: this may prompt us to consider household consumption at a more disaggregated level.

... but better results were obtained for expenditure on certain goods or services

For forecasting the consumption of different goods, there are already reliable short-term indicators for estimating, virtually in real time, automobile purchases and expenditure on energy. Priority was therefore given to three other items, which were targeted to test the performance of Google Trends data: expenditure on food, clothing and home equipment goods. Results at this level of detail were more positive. Nevertheless, for these three items, the forecasting performances of Google Trends series were not equivalent.

Google Trends improved forecasts for expenditure on clothing and home equipment

For expenditure on clothing, the explanatory variables that were seen most often in the Bayesian approach were the first two time lags in the dependant variable, illustrating the high auto-correlation of the series. The Google Trends categories "Apparel" and "Sporting goods", topics that are clearly linked with the type of expenditure being studied, also appeared among the most probable regressors.

Box 2 - Bayesian variable selection and combination of models

The Bayesian approach (see, for example *Raftery et al. 1997*) consists in fixing a prior probability on the parameters of the model (which in this case includes the composition of the model, i.e. the fact of including one variable or another), and deducing from this a posterior probability, given the observations that are available. If Y denotes the observations of a variable of interest, and X the observations of available K regressors, and M a model defined by its regressors and modelling parameters, then a posterior probability is obtained using the Bayes formula based on a prior probability $P(M_i)$ and on the marginal likelihood of the model $P(Y|M_i)$:

$$P(M_i|Y)=\frac{P(Y|M_i)P(M_i)}{P(Y|X)}=\frac{P(Y|M_i)P(M_i)}{\sum_{l=1}^{2^k}P(Y|M_l)P(M_l)}$$

Combining models using a Bayesian approach involves combining forecasts from different models by weighting them according to their posterior probability. The forecast from a given variable of interest y will therefore be a combination of the forecasts obtained from the different models weighted by their posterior probability as defined below:

$$\hat{y}_{T+h}=\sum_{i=l}^{2^k}P(M_i|y_1,...y_T)\hat{y}_{T+h,i}$$

The advantage of this approach lies in the fact that the modeller can fix the prior distributions of the models and in this way favour parsimonious models, by assigning a smaller weight to those which include a large number of regressors. Indeed, when a large number of variables are available, there is a risk of ending up with a model that is good at accounting for observations from the past, but which is not very good at predicting (we then talk of "overfitting", *Box 3*). Predicting is achieved by combining the L most likely models. The modeller decides the number of models to be selected, taking into account the posterior distribution of the models: thus, if the distribution is flat, it is preferable to select a fairly large number L of models since they are equiprobable, whereas a smaller number would be selected if only a few models differed significantly from the rest. The prediction can be written as follows:

$$\hat{y}_{T+h}=\frac{\sum_{i=l}^{L}P(M_i|y_1,...,y_T)\hat{Y}_{T+h,i}}{\sum_{i=l}^{L}P(M_i|y_1,...,y_T)}$$

Thus, some variables were selected more often in combined models than others: we therefore carried out indirect variable selection.■

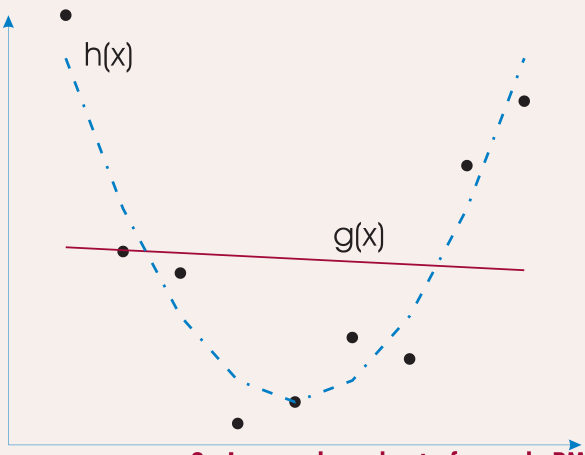
Box 3 - In forecasting, it is sometimes best to leave well alone

Why should we try to promote parsimonious models when we already have a large number of variables that all seem perfectly relevant? Introducing a large number of variables can always improve the fit of the models to the observations, but it can also diminish the predictive power of these models. A model which fits a given period perfectly will be difficult to generalise; in other words, it could become much less reliable if new observations were introduced.

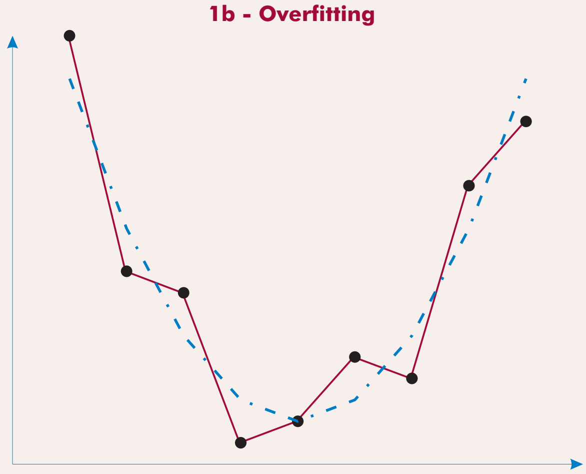
To take this into account, we consider data generated from a function h and an error term ε: $y(x)=h(x)+\varepsilon$. In practice, and as illustrated in this very simple example, the dependence of y on x is quadratic. The forecaster, who in principle is unaware of this "true" relationship, may look for the simplest type of model g(x) which consists of a dependence that is simply linear. This model introduces few parameters but generates considerable bias, wich boils down to a very poor fit (*Graph 1a*). Conversely, by adding a large number of parameters into the model, the forecaster can produce a perfect fit (*Graph 1b*). However, this perfect fit wrongly captures the error term that intervenes here in generating these data, which is by definition unpredictable. To avoid the "overfitting" shown in the right-hand graph, but reduce the bias that can be seen in the left-hand graph, there has to be a bias-variance trade-off. Usually, performance criteria assessed out-of-sample are used to judge this relevance, such as the RMSE (i.e. the root mean square error between forecast and the actual outcome).

This aspect is not automatically taken into account in the Bayesian approach, as the posterior probability of a model is based on the likelihood of the observations and not on a predictive performance criterion. The predictive performance of models depends crucially on the average size of the prior models (*via* the prior distribution of the selected models). To select the best combination of models according to their predictive performances, and to make a bias-variance trade-off, we carry out a cross-validation by calculating the out-of-sample RMSE for the models selected by the algorithm for different values of average prior size of the models. The overfitting phenomenon occurs when the prior size of the model increases: when models with a large number of variables are preferred, the in-sample RMSE improves constantly, whereas the out-of-sample RMSE eventually deteriorates (*Graph 2*). The value that is finally selected for the prior size of the model will be the one that minimises the out-of-sample RMSE. ■

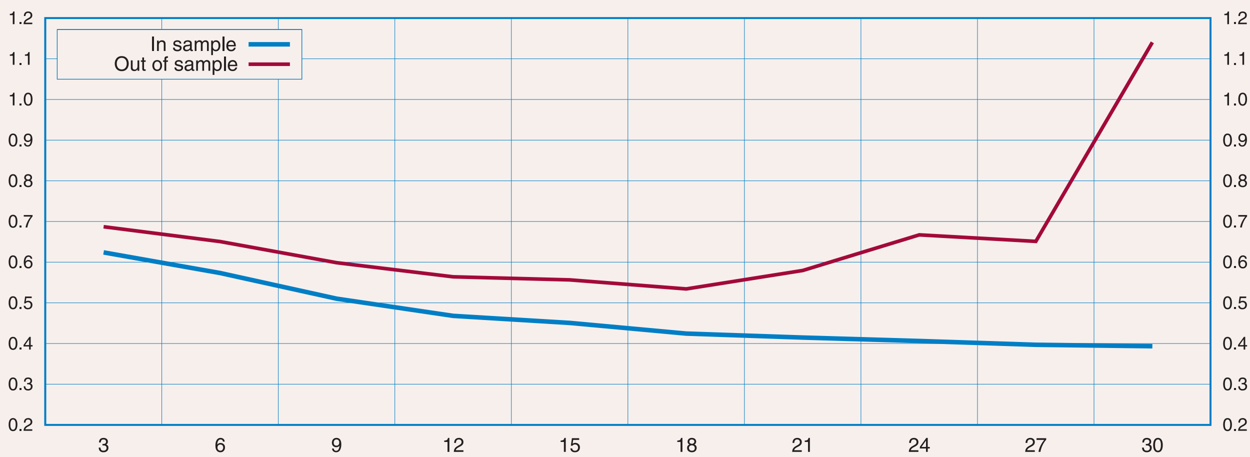
1a - Linear estimate



1b - Overfitting



2 - In-sample and out-of-sample RMSE as a function of expected prior model size



By using Google Trends categories, forecasting this item of consumption expenditure was improved: the out-of-sample root mean square error (RMSE) was about 10% less than for a simple autoregressive model (*Box 4*). The results for forecasting purchases of home household durables were also positive: the explanatory variable that emerged most often was the category "Home furnishings", where once again the topic is closely linked with the type of purchase being considered. However, this improvement in forecasting was more modest, since the RMSE was reduced by less than 5% compared with a simple model using only the dynamic of the variable to be predicted (ARMA).

The contribution was less relevant for expenditure on food...

Finally, concerning expenditure on food, using the Google trending searches did not improve forecasting when compared to an autoregressive model. However, the Google Trends categories "Tobacco products" and "Alcoholic beverages" appeared alongside the first time lag of the modelled variable among the most frequently selected regressors. The first time lag in the "Sports"category was also a frequently seen regressor, although in this case it was more difficult to establish a direct link with expenditure on food.

... and on transport services

Concerning the consumption of services, the most conclusive results were those obtained for "Transport". Although using the Google Trends categories did not improve the quality of "out-of-sample" forecasting, some of the categories appeared among the most probable regressors in the Bayesian approach, such as "Air Travel" and "Hotels and Accommodations". In-sample, "Air Travel" provided a good fit to the uneven changes in consumption expenditure on transport services in 2010, the year when air traffic was particularly disrupted in France due to the eruption of the Icelandic volcano Eyjafjallajökull in the spring and the shortage of de-icing fluid in airports during the extreme snow conditions in December.

Box 4 - Forecasting consumption expenditure on clothing can be enriched by including Google Trends categories, using the Bayesian approach

For some types of purchase, such as spending on clothing, using the Bayesian approach and including Google Trends categories can improve forecasting by expanding it, in particular compared to a simple model that uses only the past dynamics of the series (ARMA type). In addition, this approach has two interesting features:

- Posterior probabilities for the different possible models are certainly fairly similar, which is a good reason to try and combine them. In this way, predictive performances can be improved (*Graph 1*).
- By parameterising expected prior model size, the model can be optimised according to these predictive performances; in this way the risk of overfitting can be controlled (*Box 3*).

The Bayesian approach can be interpreted as indirect variable selection, *via* the selection of models with a high posterior probability, as it determines the regressors that have the highest probability of being included in the selected models. The Bayesian approach is equivalent overall in terms of performance to variable selection methods using traditional iterative algorithms (Stepwise or Pc-Gets algorithm).

The most probable Google Trends categories obtained using the Bayesian approach for predicting clothing expenditure are listed in the following table.

In this exercise, the first two time lags for the dependant variable emerge most often, which shows that the series is very auto-correlated; in particular, after a surge in spending associated with an exceptional period of sales, for example, the probability of a reaction the following month is high. Two Google Trends categories that are directly linked with expenditure on clothing also appear among the most probable explanatory variables: these are the categories "Apparel" and "Sporting goods". However, there is no guarantee as to the relevance of all the series selected, as some may be strongly correlated with clothing expenditure, but with no clear causal link (e.g. "Books and Literature" or "Vehicle shopping").

If we retain only the categories that are on face value the most relevant out of those categories that are the most probable, we can improve the monthly forecasts while at the same time favouring the simplicity of the model and avoiding certain pitfalls such as problems of multicolinearity. Thus by including Google Trends categories ("Sporting goods" and "Apparel") the forecast for expenditure on clothing can be improved in relation to an autoregressive model using only the dynamics of the series: the out-of-sample RMSE is reduced by 11% (*Graph 2*). ■

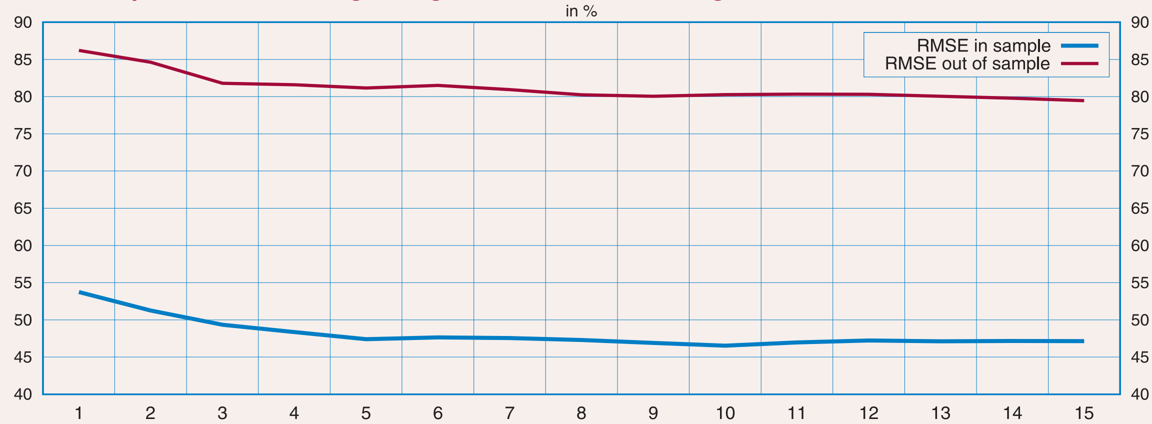
Input from Google Trends for forecasting the short-term economic outlook in France: limited avenues

Probability of inclusion and mean coefficient of the most probable explanatory variables for growth in expenditure on clothing

Name of the regressor	Posterior probability of inclusion	Mean coefficient
First time lag of the modelled variable	1.00	-0.56 (0.10)
Second time lag of the modelled variable	0.99	-0.36 (0.11)
Sport (category Google Trends)	0.96	-0.18 (0.06)
Apparel (category Google Trends)	0.95	0.39 (0.15)
Books and Literature(category Google Trends. First time lag)	0.84	-0.25 (0.13)
Home furnishings (category Google Trends)	0.58	-0.14 (0.14)
Vehicle shopping (category Google Trends)	0.46	-0.10 (0.12)
Sport goods (category Google Trends)	0.44	0.14 (0.17)

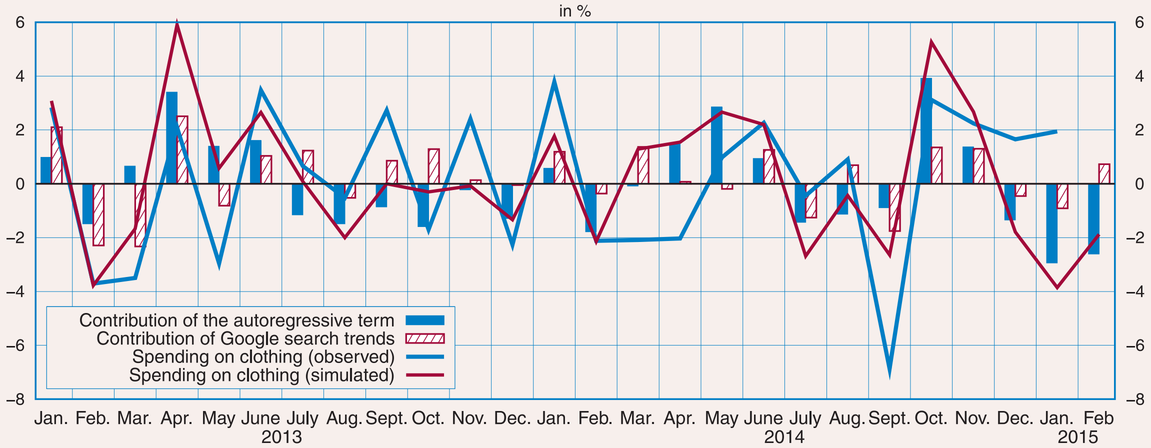
How to read it: the posterior probability of inclusion in the model corresponds to the sum of the posterior probabilities of the models in which the regressor appears. This is 95% for the Google Trends category "Apparel". The mean coefficient corresponds to the mean of the coefficient on the regressor for the most probable models, weighted by the posterior probability of the model. The standard errors of the means are given in brackets.

1 - Expenditure on clothing: change in mean error according to the number of models selected



How to read it: Mean square errors expressed as a percentage of the variance of the series. When five models are selected, mean errors in-sample and out-of-sample are 47% and 81% respectively of the variance of the series.

2 - Monthly change in clothing consumption expenditure



How to read it: the historic series of expenditure on clothing was compared from the beginning of 2013 with a simple model using the first time lag in the series and the Google trending searches "Apparel" and "Sporting goods" (t-stats given in brackets):contribution of the autoregressive term

$$\text{Spending\_clothing}_t = \underset{(-6,4)}{-0,62}.\ \text{Spending\_clothing}_{t-1}\ \underset{(-5,1)}{-0,51}.\ \text{Spending\_clothing}_{t-2}\ \underset{(-2,8)}{-0,24}.\ \text{Spending\_clothing}_{t-3}$$
$$\underset{(2,5)}{+0,23}\ \text{Google\_sport\_goods}_t\ \underset{(3,8)}{+0,42}\ \text{Google\_apparel}_t$$

R2 ajusted = 0.55 RMSE (in sample) = 2.21 RMSE (out of sample) = 2.80

Conclusion: Internet searches are informative but in practice prove to be somewhat limited for short-term forecasting

From the different types of modelling tested, it appears that adding Google trending searches does not improve the forecasting of monthly household expenditure except in specifically targeted cases. More precisely, these series do not improve forecasts of monthly household consumption expenditure on goods or services when they are considered at an aggregated level, due to the very wide variety of changes per product. On the other hand, results for the purchase of certain goods (especially clothing and household durables) are more positive and some Google Trends categories do indeed seem to be probable explanatory variables. However, when there is any improvement in forecasting, it is always small. In addition, the fact that the forecast is improved for only a few products suggests that there is a risk that these favourable results are down to "chance": it may be non-zero, but nevertheless this risk seems small since the most probable Google Trends categories for modelling consumption of the products studied are usually directly linked to these products.

As shown by the example of the Google Flu indicator, which is built on the same principle, the long-standing validity of these results must be regularly checked. The stability of their method of construction over time would seem to be difficult to guarantee, subject as it is to strategic or technological changes in Google and in the search engine, and also to the behaviour of Internet users.

These limitations would be even more significant if Google Trends were to be used to produce consumer statistics.■

Bibliography

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

Choi H. et Varian H. (2009), *Predicting the Present with Google Trends,* Technical report, Google.

Faure M.-E., Kerdrain C. et Soual H. (2012), "La consommation des ménages dans la crise", *Note de Conjoncture*, juin, p. 23-37, INSEE.

Ginsberg J., Mohebbi M. H., Patel R. S., Brammer L., Smolinski M. S. et Brilliant L. (2009), "Detecting influenza epidemics using search engine query data", *Nature 457*, p. 1012-1014.

Kulkarni R., Haynes K., Stough R. et Paelinck J. (2009),"Forecasting housing prices with Google econometrics", *Research Paper* 2009-10, George Mason University School of Public Policy.

Krankadler É. (2014), "Où fait-on ses courses ?", *Insee Première*, n° 1526, INSEE.

Lazer D., Kennedy R,. King G. et Vespignani A. (2014), "The parable of Google Flu: traps in big data analysis", *Science*, 343, p. 1203-1205.

Raftery A.E, Madigan D. et Hoeting J.A (1997), "Bayesian Model Averaging for Linear Regression Models", *Journal of the American Statistical Association*, Vol. 92, n° 437, p. 179-191.

Vosen S. et Schmidt T. (2011), "Forecasting Private Consumption: Survey-Based Indicators vs. Google Trends", *Journal of Forecasting*, Vol. 30, n° 6, p. 565-578.■