Unemployment? Google it!

Analyzing the usability of Google queries in order to predict unemployment

Gijs te Brake
MSc Economics, Universitat de Barcelona

Advisor: Raul Ramos*
AQR-IREA, Universitat de Barcelona

June 20, 2017

Abstract

During the last years the accessibility of big data has risen exponentially mainly due to the increase of internet usage. The biggest internet search engine Google Sites made statistics about the search queries public in real-time. In this paper these search queries are exploited in order to analyze whether this new type of data have the capability to improve the traditional econometric forecasting models. More precisely, this paper analysis the usability of Google search terms in order to forecast the unemployment rate in the Netherlands. This is done by creating a variable based on the volume of search terms submitted on Google (Google Indicator). The predictive capacity of the Google Indicator is measured by comparing the accuracy of a benchmark model versus an augmented model where the Google Indicator is added. The findings show that the Google augmented models produce up to 27.8% more accurate forecasts when considering a one-month ahead forecast horizon. During more recent sub-periods this improvement is even higher, reaching forecast performances that are 34.6% more accurate. However, the predictive power of the Google Indicator is diminishing when the forecast period is extended. This indicates that the use of Google data is mainly beneficial for short-term predictions.

Keywords: *Big Data, Unemployment, Google Indicator, Forecasting, Nowcasting*

JEL Classification: J64, C53, C55

---

1 Introduction

In recent years new methods of economic research aroused due to the accessibility of big data. Big data is a broad term to describe data sets which are characterized by such a high volume, velocity and variety, that they require specific technology and analytical methods for its transformation into value (Mauro et al., 2016). The data sets are growing rapidly. The same amount of data which was created until 2003 is now created every two days (Kuhn and Mansour, 2014). The rise of big data is creating new possibilities. Exploiting big data within the economic field creates the possibility to measure previously unmeasurable activities (Varian, 2010). More precisely, big data can be used to improve the nowcasting and forecasting models for economic indicators (Falorsi et al., 2015).

Various authors like O'Leary (2013) and Breur (2015) highlighted the relevance of internet usage in the development of big data. According to their findings is the internet the main reason for the rapid growth of big data. Online activity has become an inherent part of modern society. The internet has become a major source of information (Pavlicek and Kristoufek, 2015). According to a publication from Eurostat[1] in 2016 85% of the EU-28 households internet have access. This is 30% higher than in 2007. The highest proportion (97%) of households with internet access in 2016 was recorded in Luxembourg and in the Netherlands. Furthermore, 82% of all individuals, aged between 16-74 years, used the internet (at least once within the three months prior to the survey date). This number is even higher in countries such as the United Kingdom, Finland, the Netherlands and Germany. In these countries at least 9 out of 10 individuals used the internet. More than two thirds (71%) of individuals in the EU-28 accessed the internet on a daily basis. The internet is widely used in order to participate in social networking, buy products and services and gather information Eurostat (2017). Due to the intense use of internet by society it is as well used a lot in order to search for jobs in a variety of ways. The internet is becoming an essential job-search tool. Online job search (OJS) have been studied by Kuhn (2004, 2014). His results are indicating that OJS significantly increased job seekers changes of finding work and that during the last decade the internet became increasingly more important for the labor market. Nowadays there are a variety of well-known, high-traffic websites devoted to job search (for example, Monsterboard and Randstad). Furthermore, according to more recent research, OJS has become the preferred method of search for nearly all types of job seekers (Faberman and Kudlyak, 2016). Recently the European Union embraced the internet in order to stimulate the matching process between job seekers and jobs across European borders. The online platform EURES is mediating in the job market.[2]

[1]Source: Eurostat. Accessed: 2 April 2017.
http://ec.europa.eu/eurostat/statistics-explained/index.php/Digital_economy_and_society_statistics_-_households_and_individuals

[2]Source: Eures, 2017 - The EURES online portal contains job vacancies, enabling jobseekers and employers to search and launch intra-EU recruitments. The EURES is stimulating online job searches in order to resolve the mismatch in local labour markets between supply and demand. Accessed: 2 April 2017.
https://ec.europa.eu/eures/public/homepage

The main entrance of internet are the search engines (Broder, 2002). Google Sites is the worldwide leading search engine with 89 percent of search queries submitted. This equals over a 100 billion monthly searches on Google every month.[3] The two other main search engines are Bing and Yahoo! with a market share of 4,6% and 3,1% respectively.[4]

Google Sites introduced in 2007 the application Google Trends. This application is collecting the submitted search queries and it is publicly available. The data of the conducted searches are available from 2004 on. Google's dataset is because of the trillions searches that take place every year one of the world's largest real time datasets. The Google Trends application allows to measure the interest in a particular topic or search term over time for a specific location around the world.[5]

One of the main advantages of the Google Trends data is that it becomes available in real-time (Choi and Varian, 2012). The real-time data sources creates the possibility to assign real-time macroeconomic activity. Traditionally, the data corresponding with economic time-series like prices, inflation and unemployment rate are published a certain time after the period to which they refer. However, in many situations it is useful to obtain reliable indicators of the ongoing process before the publication of the traditional data. Another word for this phenomena is *nowcasting*; nowcasting uses currently available data to provide timely estimates of macroeconomic variables weeks or even months before their initial estimates are produced (Koop and Onorante, 2013). After the insights provided in the pioneering papers of Choi and Varian (2009a, 2009b), the benefits and desirability of nowcasting became more clear. This resulted in an expanding literature describing the usability of the real-time data provided by Google Trends.

The usefulness of the data from the Google Trends application is exploited in different economic areas. For example in the financial field, it is critical for banks to obtain real-time information in order to make better decisions (Aruoba and Diebold, 2010). Recently, several banks analyzed the Google Trends data to obtain information about the current state of the inflation rate, the housing prices and the unemployment rate. The Central Bank of England (McLaren and Shanbhogue, 2011), Israel (Suhoy, 2009), Turkey (Chadwick and Sengül, 2012) and Italy (D'Amuri and Marcucci, 2012) explored Google Trends in order to obtain insights in the economic indicators. Google Trends data is as well used in the stock market. Bijl et al. (2016), Hamid and Heiden (2015) and Preis et al. (2013) measured the change in search volume for specific search words relating to the American stock market as a proxy for the investors attention. Their findings are contradictory. According to the paper of Hamid and Heiden (2015) is an increase in the search terms a sign for an uplifting stock market. On the contrary, the findings of Bijl et al. (2016) and Preis et al. (2013) conclude that high levels of Google search terms predict low future stock values. Preis et al. (2013) conclude that an increase in Google search terms can be interpreted as an "early warning sign". The mixed findings are probably due to the different time period of analyze. Despite the difference in the outcome variable, the

[3]Source: Google Internal Data. Accessed: 8 April 2017.
https://research.google.com/pubs/DataManagement.html

[4]Source: Statista - The Statistics Portal. Accessed: 8 April 2017.
https://www.statista.com/statistics/216573/worldwide-market-share-of-search-engines/

[5]Source: Google NewsLab, 2016. Accessed: 10 April 2017.
https://medium.com/google-news-lab/what-is-google-trends-data-and-what-does-it-mean-b48f07342ee8

papers agree in the predictive power of Google search terms. The predictive power of the change in search volume is improving the models especially during a short time horizon. The usability of including Google Trends data diminishes after approximately four weeks.

Rivera (2016) used Google queries in order to forecast the number of hotel non-resident registrations in Puerto Rico. In his paper he utilizes nine different search queries, among them are; "Puerto Rico hotels", "Puerto Rico flights" and "Puerto Rico vacation". His paper finds evidence that Google search terms are a reliable indicator of future hotel registrations. However, the beneficial performances of Google queries are becoming feasible after a horizon of six months. This indicates that using Google queries is not improving the performance of the models in the short term, most likely because people tend to book their holiday in advance.

Veldhuizen (2016) analyzed the correlation between Google search data and transactions on the Dutch housing market. He uses Google search terms conducted in the Netherlands for the native translation of the word 'mortgage' to predict monthly housing transactions. The authors find evidence that the conducted Google queries six and nine months prior to the current month are associated with the actual amount of transactions in the current month. They conclude that adding Google search data increases the explanatory power of their model to predict housing transactions. Their results indicate that policy-makers can use Google search data to track and forecast the development in the Dutch housing market.

Recently, Google Trends have been exploited in other fields besides economics. There are various papers in epidemiology which are analyzing the correlation between Google search behaviour and epidemic outbreaks. Before the possibility to exploit Google Trends data, the models relied on traditional indicators like; clinical symptoms, virology laboratory results, hospital admission (Dugas et al., 2013). The process of data collecting and processing causes a 1-2 week reporting lag. By using Google search data this lag can be reduced till one day. According to the paper of Dugas et al. (2013) is there a high correlation between online search behaviour and an epidemic outbreak in the near future. Their findings indicates that models based on Google search data outperforms the traditional models. This result is recently confirmed by Teng (2017). His research was focused on the ZIKV epidemy and his result showed that the Zika-related google terms had a strong correlation with confirmed, suspected and total cases of ZIKV. Besides improving the nowcasting accuracy, he stated that the forecasting ability of a model based on Google search is as well higher.

The different applications in multiple fields are indicating that exploiting this new kind of data tends to improve the existing time-series models. However, the predictive power of the Google search terms depends on the variable of interest. This is due to the specific character of the underlying research. The stock market is characterized by its dynamic character. The use of Google search terms is therefore mainly beneficial in the short term, whereas the benefit of the search data in the case of the housing market becomes feasible after a longer period.

Within the broad field wherein Google data is utilized, this paper analyzes Google search behaviour and the labour market. This paper is inspired by the findings of Choi and Varian's (2009a) pioneering paper. In their paper they motivated the usability of Google search terms in order to predict the real-time unemployment rate in the United States. This paper is focusing on the Netherlands instead. To my knowledge is this the first approach which uses this country as a case-study. Besides contributing with a new case study is this paper analyzing whether models based on Google search terms are especially useful during specific periods of the business cycle. Furthermore, instead of relying only on the Google Trends application is the Google AdWords applicaction also used.

The rest of the paper is organized as follow. Chapter 2 provides an overview of the existing literature. Chapter 3 describes the data used to predict the Dutch unemployment rate, with a particular emphasis on the Google Indicator. In Chapter 4 the empirical methods are explained while Chapter 5 presents the corresponding results. In Chapter 6 the sensitivity and the robustness of the empirical results are explored and, finally, Chapter 7 contains a discussion and the conclusions of this research.

2 Literature review

The use of internet is relatively new and until recent years little studied. Therefore, this section analyzes the existing literature.

The first analysis addressing the usability of Internet is performed by Ettredge et al. (2005). They pointed out that search engines provide rapid and easy access to websites and information sources. The authors stated that the search engines contain useful information about people's interest, concerns and needs. Their study had an experimental character due to the fact that internet was not widely used at that time and the accessibility of internet data before the introduction of Google Trends was limited. Nevertheless, their preliminary results indicates a significant association between (un)employment-related searches and the official unemployment rates in the United States.

Choi and Varian (2009a, 2009b) continue in this line of research by analyzing how internet search data, more precisely, Google Trends data can be exploited in order to improve the prediction accuracy of the unemployment forecasting models. Whereas Ettredge et al. (2005) were forced to rely on limited data, Choi and Varian (2009a) had the disposal of a more complete dataset provided by the Google Trends application. In their seminal papers they introduce the term *"predicting the present"* in order to describe the usability of the Google Trends application. In their papers they focus on the labour market of the United States and their approach included an augmentation of a standard time-series forecasting model with the Google Trends series. They constructed the Google Indicator by analyzing the search intensity of the words falling under the Google category "Local/Jobs" and "Society/Social Service/Welfare & Unemployment". They demonstrate that the model augmented with the Google Indicator outperforms the standard models.

The papers of Choi and Varian (2009a, 2009b) together with the introduction of Google Trends stimulated a lot of the recent papers in this field. Since then, Google Trends data has been used to forecast unemployment in different countries. In particular; Belgium (Bughin, 2011), Brazil (Lasso and Snijders, 2016), China (Su, 2014), Germany (Askitas and Zimmerman, 2009), Israel (Suhoy, 2009), Italy (D'Amuri, 2009), (D'Amuri and Marcucci, 2012) and (Falorsi et al., 2015), Norway (Anvik and Gjelstad, 2010), Spain (Vicente et al., 2015), Turkey (Chadwick and Sengül, 2012), the United Kingdom (McLaren and Shanbhogue, 2011), The United States (D'Amuri and Marcucci, 2010), (Choi and Varian, 2012) and (Tuhkuri, 2015) and finally, the Visegrad group countries (Pavlicek and Kristoufek, 2015).

The methodology applied in most of the existing literature in this field is similar. Researchers proceed mainly in the following three steps: Construction of the Google Indicator, deciding the appropriate benchmark model and finally, a comparison of the forecasting capacity between the benchmark and the Google augmented model.

The choice of search terms is crucial for the studies in this field. There are different strategies in order to obtain the Google Indicator. Broadly, these strategies can be divided into three different groups; using predefined Google categories, using multiple words in order to define the search terms correlating with unemployment or the use of a single word.

Google Trends automatically classifies the conducted search terms into thirty different categories. Within these categories Google Trends classifies the search terms in a more detailed way in 250 different subcategories. As mentioned before, the use of Google categories is performed firstly by Choi and Varian (2009a). Suhoy (2009) and Bughin (2011) exploited the Google categories as well by using the subcategory "Local/Jobs". By using the Google Trends categories the researchers are obtaining information regarding the search intensity of a specific Google Trends category in comparison to other categories. This strategy was mostly used during the first stage of research within this field. According to Choi and Varian (2009a) is the underlying data by using search categories the highest due to the fact that a wide range of search queries are classified under a certain subcategory. On the other hand, within the range of a category are not all words directly correlated with unemployment searches this can introduce an unnecessarily amount of noise. For example, a search query as; "Desk chair" falls under the before mentioned category "Local/Jobs", however this search query is not directly associated with unemployment.

The most performed strategy in order to obtain the Google Indicator is the use of a single word. The specific word in order to obtain the Google indicator is similar in all existing works: "*Jobs*". For example D'amuri (2009, 2010, 2012), Fondeur and Karamé (2013), Su (2014), Vincente et al. (2015), Pavlicek and Kristoufek (2015) and Falorsi (2015) used native words for "jobs", "job offers" or "work". According to D'amuri and Marcucci (2010), the search terms "jobs" has the highest search volume and they belief that it is used across the broadest range of job seekers. They discuss the possibility of including in the model other, less frequently used, job-search-related search terms in order to increase the search volume underlying the Google Indicator, but according to them can the inclusion of other words bias the Google Indicator and its ability to predict the unemployment rate. Fondeur and Karamé (2013) agree with this strategy and they as well use the single term "jobs". They point out that this search term is directly connected with job searches because it is the simplest way to find the websites where the jobs are posted.

Other studies used more sophisticated ways to obtain their Google Indicator. These works are exploiting different search terms in order to obtain a Google Indicator which proxies the job search on the labour market. The first paper following this strategy is Askitas and Zimmerman (2009). They use a set of eight keywords in order to "weed out the noisy activity and get to the signal in any kind of effective way". They use a Boolean Operator in order to group the keywords together in four different groups and thereafter they analyze which words are correlated with the monthly evolution of the actual unemployment rate. The first group represents the words relating to "unemployment office". The second group is composed by the word "unemployment rate". The third group consists out of the search terms which are expected to relate with "Human Resource consulting" and the last group is corresponding with the most popular "job boards" in Germany. Anvik and Gjelstad (2010) uses the same approach as Askitas and Zimmerman (2009). Tuhkuri (2015) uses as well different search terms. Additionally, he exploited another feature of Google; *Google Adwords* in order to get insights in the underlying volume (number of monthly searches submitted for a certain query). This is the first analysis wherein the actual Google search volumes is used. Furthermore, Tuhkuri (2015) decided to construct the Google Indicator solely based on words specifically related to unemployment benefits, instead of the more general term “jobs”. According to Tuhkuri (2015) displaced workers are likely to submit searches firstly to the unemployment benefits. In his approach he initially obtained 125 different search terms, eventually he used the thirteen search terms with the highest search volume. Additionally he uses the Boolean algorithm to provide different weights for the search terms depending on the search volume.

Before it is possible to analyze the usability of the Google augmented models, a benchmark model is needed in order to compare the relative performance. The strategy to construct the benchmark model is similar across the different works. Like mentioned before, Choi and Varian (2009a) were the first using Google Trends in order to improve the forecasting accuracy of the unemployment forecasting models. In their paper they used an univariate time series approach in order to build their benchmark model. The emphasis of this type of modelling is determined by using the information of the past value of a variable in order to forecast its future value(s). The studies afterwards are in line with Choi and Varian’s (2009a) paper and are using the well-known Box-Jenkins approach in order to obtain autoregressive integrated moving average (ARIMA) models as a benchmark (Box et al., 2015). In fact, ARIMA models are the most commonly used benchmark models for unemployment forecasting according to Fondeur and Karamé (2013). However, the level of sophistication of the ARIMA models varies among the papers. Many studies, including Choi and Varian (2009a), D’Amuri (2010), McLaren and Shanbhogue (2011), Choi and Varian (2012), Chadwick and Sengül (2012), Pavlicek and Kristoufek (2015), Tuhkuri (2015) and Lasso and Snijders (2016) adopt an AR (1) model as benchmark. In comparison, Suhoy (2013) uses an ARMA (2,2) model while Anvik and Gjelstad (2010) employs an AR (3) process as a benchmark. Vicente et al. (2015) considers an ARIMA (0,1,2) process in order to analyse the unemployment movements.

Bughin (2011) and Falorsi et al. (2015) use Vector Error Correction techniques (VEC models). Falorsi et al. (2015) mention that the univariate modelling techniques ignore possible cross-correlation between the unemployment and Google search term time series. This can result in a loss of information and the choice of an unsuitable model. According to them, the two before mentioned time series will provide information on one another and in the long they present the same trend. Therefore, cointegration constitutes an element of improvement of the model in terms of the predictive power with respect to the use of the ARIMA modelling techniques.

Most recently, Scott and Varian (2015) use more sophisticated strategies in other to obtain the benchmark model. They introduce the Bayesian Variable Selection and the corresponding Bayesian Structural time series (BSTS) modelling technique (Scott and Varian, 2015) In their papers they continuing using the AR (1) model as the benchmark. However, they use the BSTS framework. The BSTS modelling framework focuses specifically on search engine data which contain a large numbers of contemporaneous predictors. In comparison to the traditional framework, the BSTS is inspired by machine learning techniques. Previously, the search terms used by other researches to obtain the Google Indicator and the benchmark model were based on human judgement and economic intuition. However, there are millions of different search terms conducted on Google on a monthly base. This large amount of potential predictors can increase the chance of model misspecification. To overcome that possible problem, the BSTS is using algorithms in order to determine the causal impact of the used search terms.

After the construction of the benchmark model, an assessment is made by analyzing the forecast performance of the Google augmented model in comparison to the benchmark model. Before examining the predictive accuracy the existing studies analyse whether the Google Indicator is associated with the dependent variable, unemployment rate. Not surprisingly, every paper concludes that indeed there is a significant correlation between those variables. Thereafter, the forecasting accuracy is compared by applying different instruments and examining various forecasting time horizons. Most of the existing papers are employing the (pseudo) out-of-sample forecast framework with a rolling window. In most cases this is done by performing a one-step-ahead out-of-sample prediction in order to asses the current conditions. Among others this is the approach performed by Askitas and Zimmerman (2009), Choi and Varian (2012), Chadwick and Sengül (2012), Fondeur and Karamé (2013) and Pavlicek and Kristoufek (2015). A few attempts are undertaken to asses a longer forecasting period. For example D'Amuri and Marcucci (2012) analyze the forecasting performance in a horse-race setup of more than 500 models both at a one-, two- and three-month horizon. More recently, Tuhkuri (2015) pushed this line of work even further by assessing the near future in addition to the present. In his paper, he obtains a potential value of the unemployment rate by using Google data in order to forecast until 5 months before the present.

Most studies evaluate the usefulness of the incorporation of the Google Indicator by using the Mean Absolute Error (MAE) or the Root Mean Squared Error (RMSE). Additionally, the studies of Fondeur and Karamé (2013), Vicente et al. (2015) and Tuhkuri (2015), used the Mean Absolute Percentage Error (MAPE) as well. Other (more formal) instruments as the Diebold-Mariano test is only employed by D'Amuri and Marcucci (2012), Pavlicek and Kristoufek (2015) and Tuhkuri (2015).

The MAE measures the average magnitude of the forecast errors in the sequence. It shows the absolute differences between the actual unemployment rate and the predictions of the models. All the errors have the same weight. The expressions below show how the MAE is calculated.

$$\text{Mean Absolute Error: MAE} = m^{-1} \sum_{h=0}^{m-1} |\hat{e}_{n+h+1}| \tag{1}$$

The equation considers $n + m$ as the total observations. n denotes the observations in order to obtain the parameters of the model and m denotes the amount of observations used to analyze the forecast performance. The forecast is given by; $\hat{f}_{n+h}$ for y_{n+h+1} where h expresses a certain forecast moment up till $m - 1$. The forecast errors are denoted as $\hat{e}_{n+h+1}$. This is equal to the difference between the forecast value and the actual value: y_{n+h+1} - $\hat{f}_{n+h}$. The statistical interpretation of this type measurement tool is more intuitive in comparison to the other measurement tools because it evaluates the overall average error (Willmott and Matsuura, 2005).

The RMSE considers the square root of the average of squared errors between the actual observations and the predictions. In comparison to the MAE, the RMSE measurement emphasizes larger errors. As seen in (2) more weight is given to the larger errors because they are squared before the average is given (Chai and Draxler, 2014).

$$\text{Root mean squared error: RMSE} = \left(m^{-1} \sum_{h=0}^{m-1} \hat{e}^2_{n+h+1} \right)^{1/2} \tag{2}$$

The MAE and the RMSE have similar properties. They asses the forecast error in deviations from the independent variable and they are indifferent to the direction of the errors. Furthermore, they produce negatively-orientated outcomes, this means that the lower value is preferred. However these measures have the drawback that they are affected by the measurement unit of the independent variable. In other words, they are scale dependent and therefore they are not informing about the accuracy of our predictions. The instruments are mainly used to make comparisons between models.

Another way of measuring the forecasting errors is by using the MAPE. The instrument expresses the difference in forecast and actual observation in percentage. Similar to the MAE and the RMSE produces the MAPE as well negatively-orientated outcomes (Armstrong, 2002). This study considers a MAPE outcome of $\leq 2\%$ as a very good forecast capability. A MAPE of $< 2\% - \leq 5\%$ is considered as a good forecast capability. A model which produces a MAPE higher than 5% is seen as a model with low predictive capacity.

$$\text{Mean Absolute Percentage Error: MAPE} = m^{-1} \sum_{h=0}^{m-1} \frac{|\hat{e}_{n+h+1}|}{Y_{n+m}} \cdot 100\% \tag{3}$$

The Diebold-Mariano test (D-M test) analyzes whether the difference in forecasting accuracy between the models are statistically significant (Diebold, 2015). This is done by comparing the forecasting errors of both models with the actual unemployment series. The D-M test assesses the accuracy based on the loss differential:

$$d_t = g\{\varepsilon^2_B\}^T_t - g\{\varepsilon^2_{GI}\}^T_t \tag{4}$$

The loss differential depends on the selected loss function (g). The loss function is the sum of the squared forecast errors from respectively the benchmark and the Google augmented model. The null and alternate hypotheses are stated as:

$$H_0 : E\{\varepsilon^2_B\}^T_t = E\{\varepsilon^2_{GI}\}^T_t \tag{5}$$

$$H_A : E\{\varepsilon^2_B\}^T_t \neq E\{\varepsilon^2_{GI}\}^T_t \tag{6}$$

Under the null hypothesis there is no significant difference between the forecasting accuracy of both the models. The alternative hypothesis indicates that one model statistically outperforms the other model.

The results of the existing literature shows for almost all countries and studies an improvement in the forecasting accuracy of the Google augmented models. However, the magnitude of these improvements varies depending on the country and study. For example, Mclaren and Shanbhogue (2011) finds for England a decrease of 12%. This improvement holds for Spain where Vicente et al. (2015) obtained a decrease of 15%. The same magnitude is obtained for Brazil by Lasso (2016). Other authors like Chadwick (2012) and Fondeur and Karamé (2013) finds even a higher prediction accuracy of 27% for France and even a improvement of 48% for Turkey. A similar result is obtained by D'amuri (2009). His result indicates an improvement of 40% for Italy.

On the contrary, Choi and Varian (2012) and Tuhkuri (2015) find less modest forecast improvements for the case of the United States. The augmented model of Choi and Varian (2009a) even shows a reduction in fit in the one-step ahead out of sample forecast and Tuhkuri (2015) obtained a relatively small improvement in the MAPE (4.32%). However, these studies considers different moment in time. It turns out that the Google augmented models are especially useful in picking up turning points in the economy. For example, the augmented model used in the study of Choi and Varian (2009) shows an improvement of 13.6% during the crisis in the United States. A similar result is obtained by Tuhkuri (2015). In his paper, he reports that the prediction accuracy of the Google augmented model is approximately four times larger than the benchmark model during turning points.

Overall the existing literature shows that incorporating Google data improves the accuracy of the unemployment prediction modelling. An overview of the strategy and the results of all the before mentioned studies is included in table A.1 of the Appendix. The magnitude of those improvements differs across countries, studies and the time periods considered. However, most of the studies did not developed a more in-depth analysis with respect to the first seminal papers of Choi and Varian (2009a, 2009b). Besides D'amuri and Marcucci (2012) and the more recent work of Tuhkuri (2015) the studies are assessing solely the current situation (nowcasting) without predicting future values (forecasting). Furthermore, the majority of the existing studies are analyzing the performance of the Google augmented models considering one time period. They are not pointing out whether the obtained improvements are stable over time or whether it is time specific. It is possible that the improvements of the Google augmented model are driven by time specific moments.

This paper addresses some of the above mentioned limitations. The performance of Google augmented model is analyzed during different time periods in order to explore if the forecasting improvement is stable or whether it varies over time. Additionally to the nowcast performance, the forecast performance of the Google augmented models is analyzed considering an extended forecast horizon. Furthermore, this paper provides a novel and interesting case study. In particular, as mentioned in the Introduction, the Netherlands is one of the countries with the highest internet penetration (97%). The time period of analysis is also relevant. The introduction of Google Trends application and the possibility to analyse this data is a relatively new phenomena. Therefore, the first studies worked with a substantial shorter time period when compared with the time period considered in this study.

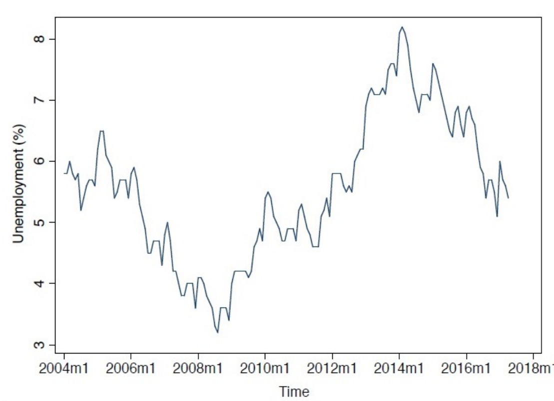
3 Data

This study combines variables from several data sources. The variable of interest is the registered unemployment rate in the Netherlands. This information is provided by the *Arbeidsdeelname en werkloosheid per maand* (monthly labor participation and unemployment survey). This is the official administrative data published on a monthly basis. The data is retrieved from the Bureau of Statistics of the Netherlands (CBS).[6] The Google Indicator is obtained by exploiting sources provided by *Google Inc.* Two applications are employed in order to obtain the data and construct a well developed Google indicator. The primary source is the Google Trends database. In addition, Google Adwords is used.

3.1 Unemployment

The data collection in order to obtain the unemployment rate in the Netherlands is performed by the CBS. The CBS publishes the unemployment rates with a three week delay with respect to the end of the month of reference. The definition of unemployment applied by the CBS is in line with the guidelines of the International Labour Organization (ILO); unemployed comprise all persons in the range of 15-75 years who are out of work, want a job, have actively sought work in the previous four weeks and are available to start work within the next fortnight. Figure 1 shows the monthly unemployment rate in the Netherlands from January 2004 until April 2017. The graph shows that the unemployment rate was at the lowest point during the beginning of 2008. At that moment were there approximately 310.000 persons out of a job. In the following six years the unemployment rate started to increase until it reaches the highest point (8,4%) during April 2014. This corresponds to 700.000 unemployment persons. According to the CBS, the financial crisis in the Netherlands was the main reason for the rise of the unemployment rate.[7]

Figure 1: Monthly evolution of the registered unemployment rate in the Netherlands. source: CBS Netherlands.



[6]Source: CBS - The unemployment data is retrieved on 25 April 2017.
http://statline.cbs.nl/Statweb/search/?Q=werkloosheid&LA=NL

[7]Source: CBS - Explanation of the unemployment evolution. Accessed: 25 April 2017.
https://www.cbs.nl/en-gb/news/2016/16/unemployment-eased-in-march

In this study the seasonal unadjusted unemployment data is used. This is in line with previous studies from Choi and Varian (2012) and Tuhkuri (2015). The analysis of this study is focused on short-term forecasting and seasonally adjusted data is partially based on forecasts. Furthermore, seasonal adjusted data is subject to larger revisions than non-seasonally adjusted data.

3.2 Google

This section, which covers Google data, consist out of two parts. First, an explanation about the Google Trends database is provided and the way of extracting the data out of that database is explained. Thereafter, a description of the strategy of selecting the right search terms in order to obtain the Google Indicator follows.

3.2.1 Google Trends application

The Google Trends application contains data on internet search volumes. The data is freely available.[8] The application measures the volume of Google searches and it compares the popularity of search terms with each other. The comparison can be narrowed to the country and even regional level and to a specific time period. The Google Trends search data are available from 2004 on. The search queries are reported as an index. Google Trends does not provide the precise number of search queries made with a specific keyword.

The index is calculated by dividing the number of searches made for a certain query term by the total number of online search queries submitted. The fraction is normalised by a scale from 0 - 100. The maximum amount of search queries for a specific search term is set equal to 100. The rest of the series is scaled according to this peak value. The normalization is an important feature of the Google data. Without this normalization it becomes more difficult to make inferences based on the data. The normalization controls for the upward trend in Google searches. In the beginning of the observation period; January 2004, was the search volume much smaller than it is during the last years. The raw search numbers would not allow to compare searches made in 2004 with more recent searches.[9] The expression below indicates how the Google Trends application operates.

$$\text{Google Trends: Search Intensity } (I) = \left[\frac{\frac{K_{t,i}}{G_{t,i}}}{\max_t \left(\frac{K_{t,i}}{G_{t,i}} \right)} \right] \cdot 100 \tag{7}$$

where: $K =$ amount of searches for a specific keyword

$G =$ Total amount of Google searches

$i,t =$ refers to a specific time and country or region

[8]Source: Google Trends. The Google data is retrieved during 14 consecutive day between 1-15 May 2017 https://trends.google.com/trends/

[9]source: Google Internal data. Accessed: 2 May 2017. https://medium.com/google-news-lab/what-is-google-trends-data-and-what-does-it-mean-b48f07342ee8

As seen in (7), the search intensity for a specific keyword denoted by I depends on the amount of searches for this keyword (K) in comparison to the total amount of Google searches (G) given a certain time and region. In other words, the Index is constructed as a proportion of the total amount of searches conducted in the given region during the same time. This proportion is scaled, according to the peak value, over the observation period.

3.2.2 Google Indicator

Deciding which search queries to consider is a crucial element of using internet search data. As mentioned in the literature review, there are several strategies in order to obtain the Google Indicator. This studies obtains the Google Indicator by using multiple search terms related to unemployment. This strategy follows the reasoning of the study of Tuhkuri (2015). In his research he uses several words related to unemployment benefits. As discussed in the Literature section, most of the existing literature uses the search term “jobs” or a group of similar search terms related to the labour market. However, this studies considers search terms specifically related to unemployment. As mentioned in the study of Tuhkuri (2015) unemployed people are especially interested in searching for unemployment benefits. By using the word “jobs” we suspect to pick up more noise because this search term is as well used by employed people who want to change jobs instead of people who are unemployed or suspect to become unemployed.

During the first step of obtaining the Google Indicator there are 26 different search terms, related to unemployment analyzed. By using the Google AdWords[10] application we were able to narrow the amount of search terms down towards the eight search terms with the highest search volume. An overview of the words which are used is included in table 1. After deciding which search terms to use weights are applied. This is done by using the information of Google AdWords and by applying the boolean search operator “OR”. The advantage of this operator is that it give insights in the weight of each search term in comparison to the other search terms. A more detailed description and visualization of the boolean operator is included in figure A.1 of the Appendix.

Besides Tuhkuri (2015) this is the first study which considers the search volumes associated with the different search terms. However, the search volumes needs to be considered with caution because Google AdWords provides ranges of search volumes, for privacy reasons the precise number is not provided. Nevertheless, the ranges are detailed enough to obtain weights for the eight search terms. Expression (8) explains how the weights are obtained.

[10]Source: Google AdWords. Accessed: 3 May 2017.
https://adWords.google.com/intl/nl_nl/

$$\text{Google Indicator} = \left[\frac{(I_{w,t} \cdot SV_{w,t})}{\sum SV_t} \right] \tag{8}$$

$$\begin{aligned} \text{where:} \quad I &= \text{Index associated with searchterm } w \text{ at time } t \\ SV &= \text{Search volume associated with searchterm } w \text{ at time } t \\ \sum SV &= \text{total amount of searches for the eight chosen search terms} \end{aligned}$$

As seen (8), the weights are obtained by multiplying the Index with it's search volume w at time t. Thereafter, the sum is divided by the total amount of searces submited for the eight different search terms at time t.

Table 1: Search terms used to construct the Google Indicator[11]

No.	*English translation*	*Native word*	*Search volume (range)*	*Weight*
1	Institute for Employee Insurance	UWV	200.000 - 350.000	21
2	Unemployment benefit	Uitkering	80.000 - 100.000	9.5
3	Unemployment law	WW	80.000 - 100.000	9
4	Income support	Bijstand	50.000 - 100.000	7
5	Unemployed	Werkloosheid	25.000 - 50.000	3.5
6	Unemployed	Werkloos	25.000 - 50.000	3.5
7	Apply for Unemployment	WW uitkering	5.000 - 10.000	1
8	Amount of unemployment benefit	Hoogte uitkering	$<$ 5.000	0.5

[11] Besides Google AdWords are additional online Keyword tools used (https://app.wordtracker.com/ and http://keywordtool.io/) in order to get a reasonable insight in the average monthly searches submitted for the accosiated search terms. Nevertheless, it is not possible to obtain more accurate numbers and even the search volume range needs to be considered with caution.

4 Empirical methods

This section presents the methods used in order to analyze whether the Google Indicator improves the forecasting accuracy of the unemployment prediction models. The first part analyzes the relationship between the Google Indicator and the unemployment series. This part includes the descriptive statistics. Thereafter the benchmark model and the Google augmented model are introduced. Afterwards the forecasting performance of the benchmark model is compared with the augmented model. The first comparison considers a one-month ahead forecast. Thereafter, an analysis of the predictive performance of both models for different forecast horizons, up to 5 months is carried out. Finally, the performance of the models are explored using a sub-period analysis in order to see if the predictive performance are constant over time.

4.1 Descriptive analysis

Before inference can be made regarding the predictive power of Google search terms, both series are jointly analyzed. Figure 2 shows the monthly evolution of both series. Both series seems to follow a similar pattern. The Google Indicator seems more unstable between 2004 and 2006. A reason for the volatile behaviour of the Google Indicator during the beginning of the period can be found in the use of internet during that period. The performed search queries were substantially lower back then, this increases the volatility. However, the correlation coefficient indicates that the evolution of both series are similar.

Figure 2: Monthly evolution of the unemployment rate and the Google Indicator in the Netherlands between January 2004 and April 2017.

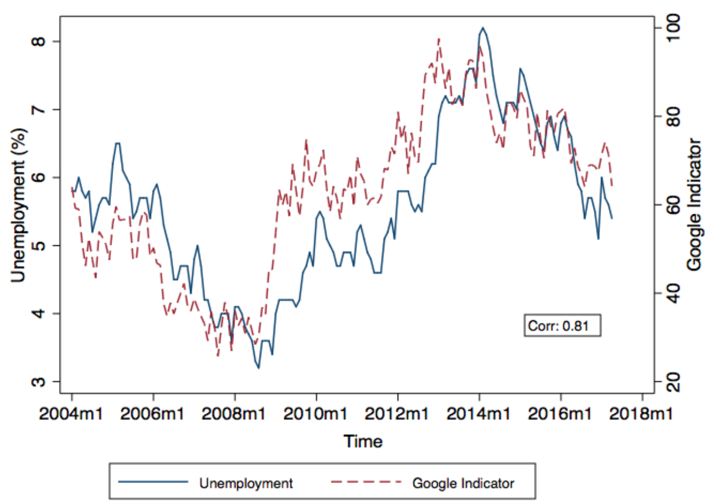


Table 2 shows the descriptive statistics for both series. As we can see, there is a higher variance for the Google Indicator in comparison to the unemployment rate. Despite that, data is not showing a strong excessive kurtosis and normality is rejected for both series.

Table 2: Descriptive statistics for the unemployment and the Google Indicator in the Netherlands

Variable	*N*	*Mean*	*Max*	*Min*	*Stand. dev.*	Skewness	Kourtois
Unemployment	160	5.55	8.2	3.2	1.17	0.40	0.01
Google Indicator	160	62.37	97.64	25.71	18.14	0.24	0.00

4.2 Forecasting models

After the descriptive analysis the benchmark- and Google augmented model are constructed. As mentioned in the Literature section, most of the existing studies use a linear nested framework wherein an univariate model of an autoregressive order 1 serves as a benchmark. This benchmark is augmented with the Google Indicator. The same approach is followed in this study.

$$\textit{Benchmark model:} \qquad U_t = \beta_0 + \beta_1 U_{t-1} + \epsilon_t \tag{9}$$

$$\textit{Google augmented model:} \qquad U_t = \beta_0 + \beta_1 U_{t-1} + \beta_2 GI_t + \epsilon_t \tag{10}$$

At first, we examine the in-sample fit for the whole period of analysis in order to analyze whether the contemporaneous value of the Google Indicator significantly improves the model. Due to the nature of the series is it reasonable to suspect that there remains a certain amount of autocorrelation in the models when an autoregressive model of order 1 is used. We account for the autocorrelation by using heteroskedasticity- and autocorrelation-consistent (HAC) standard errors developed by Newey and West (1994). Table 3 shows the estimation results of the models. The coefficient of the Google Indicator is statistically significant at a 1% level. The sign of the coefficient shows a positive relation between the unemployment rate and the Google Indicator. More precisely, a 1 percent increase of the current search intensity is causing a 0.10% increase in the unemployment rate. The R^2 shows how well the model fits the data, an increases of the R^2 is observed when the benchmark model is augmented with the Google Indicator. However the benchmark model is already explaining more then 97% of the variance. More important are the numbers associated with the Bayensian Information Criterion (BIC) and the Akaike Information Criterion (AIC). These criteria are measuring the relative quality of the statistical models, the preferred model is the one with the lower information value. The information criteria penalizes additional variables which do not significantly improve the models fit. The decrease of both information criteria suggest that the Google Indicator offers useful information in explaining the variation of the unemployment rate. The Google augmented model improves the in-sample fit. However, it is the out-of-sample performance that eventually matters for forecasting.

Table 3: Estimation results: In-sample analysis

Variable	*Benchmark model*	*Google Augmented model*
U_{t-1}	0.974^{***}	0.857^{***}
	(0.000)	(0.000)
GI		0.010^{***}
		(0.000)
Cons.	0.144	0.162^{*}
	(0.103)	(0.092)
Summary		
$\bar{R}^2$	0.948	0.959
BIC	41.1	9.3
AIC	34.9	0.1
N	159	159

The Newey and West Standard errors are given in the parentheses

$^{***}p < 0.01$ $^{**}p < 0.05$ $^{*}p < 0.1$

The evaluation period is January 2004 - April 2017.

U = Unemployment, GI = Google Indicator

4.3 Results for one step forecast

This section presents a comparison considering the out-of-sample forecast accuracy of the benchmark model versus the Google augmented model. the one-step ahead forecast is performed to analyze if the Google augmented model is able to outperform the benchmark model in "predicting the present".

The Google augmented model includes the contemporaneous Google Indicator. In other words, the Google Indicator is available in real time. The official unemployment rate is published with a three week delay. This section analyzes whether the Google searches made in real time can be used to predict the current unemployment rate.

In order to perform the one-step ahead forecast, the models are fitted on the data from the start of the period of analysis, January 2004 until December 2014 (n = 132). The forecast period is then evaluated on the series between January 2015 until April 2017 (m = 28). The one-month ahead series are generated using a rolling window. This means that after each month of prediction the model is updated with the actual value of the new observation and the first observation is dropped (Armstrong, 2002). In this way the number of the estimated coefficient stays equal during the forecasting period. Table 4 shows the one-step ahead forecast performance of both models.

The results shows that using the Google Indicator improves the model. All the proposed measures decreases when the model is augmented with the Google Indicator. Considering the MAPE, the forecasting accuracy improves with 8.7% by adding the Google Indicator. This results indicates that, considering this period of analysis, adding the Google Indicator improves the forecasting accuracy of the model.

Table 4: Evaluation statistics: One-month ahead forecast performance

Model	*MAE*	*RMSE*	*MAPE*
Benchmark model	0.251	0.310	4.02%
Google Augmented model	0.230	0.281	3.67%
Δ	8.37%	9.35%	8.70%
D-M test statistic	2.85***	n.a.	2.61***

$\Delta = The\ improvement\ in\ forecasting\ accuracy.$ ***p<0.01**p<0.05*p<0.1

The evaluation period is January 2015 - April 2017. Forecast is perfomed using a rolling window of 132 observations, starting from January 2004.

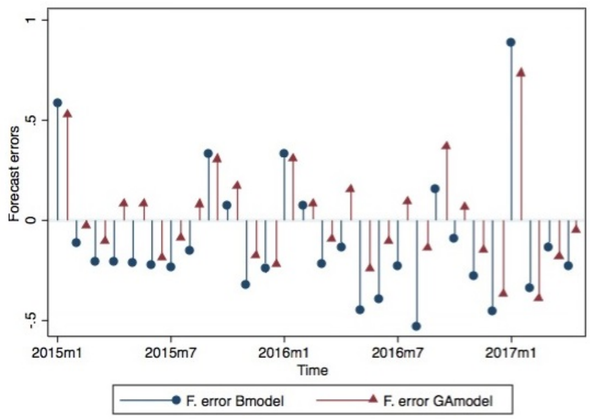
The measurement instruments shown in table 4 are constructed by taking the average of the 28 forecast periods considered. In order to analyse whether the observed difference in the values associated with those instruments are statistically significant, the D-M test is used. The obtained test statistics follows a normal asymptotically distribution (Franses, 2016). The null hypothesis can be rejected if the DM statistic falls outside the range of $-z_{\alpha/2}$ to $z_{\alpha/2}$, that is if:

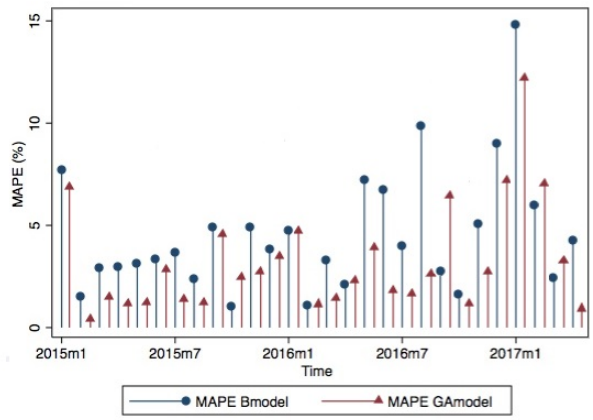
$$|DM| > z_{\alpha/2} \tag{11}$$

At a 1% significance level, the critical z-value is 2.58. The obtained test statistics exceeds the critical value and therefore the null hypothesis of no difference is rejected at a 1% significance level. The Google augmented model produces statistically significant more accurate forecasts in comparison with the benchmark model.

Figure 3 shows the forecast errors and the MAPE of both models for each produces forecast separately. The Google augmented model produces 21 times more accurate forecasts during the evaluation period which consist out of 28 periods. This means that the Google augmented model out-performance the benchmark model during 75% of the periods considered. This result is confirmed by the other measurement instruments as shown in table A.2 of the Appendix.

Figure 3: Comparison of the forecast errors and the MAPE of the benchmark model and the Google augmented model between January 2015 and April 2017.





4.4 Results for the extended forecast horizon

Whereas most of the existing literature accesses the performances of the Google Indicator in order to forecast the present (nowcasting) addresses this study as well an extended forecast horizon. Different models are constructed in order to analyze the performance of the Google Indicator over an extended forecast horizon. As shown in figure 4 are different lags of the Google Indicator included in the forecast models to examine the forecast power corresponding to different months ahead. The forecast performance of the Google augmented models are compared with the benchmark model up to 6 months ahead of the present time.

Figure 4: more-steps ahead forecast models

$$\begin{cases} \textit{Benchmark model:} & U_t = \beta_0 + \beta_1 U_{t-1} + \epsilon_t \\ \\ \textit{present:} & U_t = \beta_0 + \beta_1 U_{t-1} + \beta_1 GI + \epsilon_t \\ \textit{1 month:} & U_t = \beta_0 + \beta_1 U_{t-1} + \beta_1 GI_{t-1} + \epsilon_t \\ \textit{2 months:} & U_t = \beta_0 + \beta_1 U_{t-1} + \beta_2 GI_{t-2} + \epsilon_t \\ \textit{3 months:} & U_t = \beta_0 + \beta_1 U_{t-1} + \beta_3 GI_{t-3} + \epsilon_t \\ \textit{4 months:} & U_t = \beta_0 + \beta_1 U_{t-1} + \beta_4 GI_{t-4} + \epsilon_t \\ \textit{5 months:} & U_t = \beta_0 + \beta_1 U_{t-1} + \beta_5 GI_{t-5} + \epsilon_t \\ \textit{6 months:} & U_t = \beta_0 + \beta_1 U_{t-1} + \beta_6 GI_{t-6} + \epsilon_t \end{cases} \tag{12}$$

The forecast series are generated by using a dynamic forecast approach. This approach uses the value of the previous forecasted value of the dependent variable to compute the next one. On the other hand, the static forecast approach uses the actual values for each subsequent forecast. According to Nyberg (2010) produces dynamic models more accurate forecasts than the static models. The few studies which considered an extended time horizon, D'Amuri (2009), D'amuri and Marcucci (2012) and Tuhkuri (2015), employed a dynamic forecast approach as well. The MAPE measurement is used in order to compare the forecast performance for different time horizons ahead. This measurement is used because we are mainly interested in the forecasting accuracy of both models. As mentioned in the Literature section, the other measurement instrument are mainly used to make comparison between the models without providing insights in the actual performance.

The out-of-sample forecast performance of the extended horizon are summarized in table 5. As expected, the MAPE of both models increases over time. An extended forecast horizon increases the uncertainty and this has a negative effect on the forecasting accuracy, which results in a higher MAPE. Furthermore, the marginal predictive ability of the Google Indicator decreases between the "present period" and the one-month ahead forecast from an improvement of 27.81% till 9.71%. Nevertheless, the Google augmented model outperforms the benchmark model up to a forecast horizon of four months. A visualisation of the forecast performance of both models is included in the Appendix in Figure A.2. The monthly forecast errors and MAPE of both models are shown in Figure A.3 of the Appendix.

Table 5: MAPE performance during different forecast horizons

Model	*present*	*1 month*	*2 months*	*3 months*	*4 months*	*5 months*	*6 months*
Benchmark model	4.53%	4.53%	6.87%	7.55%	7.66%	9.36%	10.64%
Google Augmented model	3.27%	4.09%	6.07%	6.66%	6.98%	10.43%	11.65%
Δ	27.81%	9.71%	11.64%	11.79%	8.87%	-10.35%	-9.49%
D-M test statistic	3.93***	1.69*	1.65***	2.01**	1.67*	-2.39**	-2.49**

Forecast is performed using a rolling window of 48 observations, the evaluation period is January 2015 - April 2017.

$\Delta = The\ improvement\ in\ forecasting\ accuracy.$ ***p<0.01**p<0.05, *p<0.1.

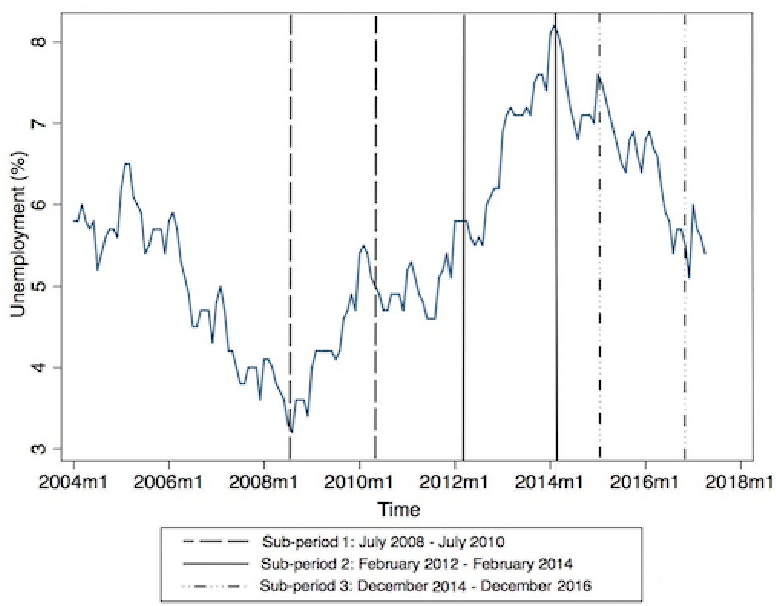
The results are indicating that the current submitted Google search queries are not only beneficial to predict the present but as well the future unemployment rate up to a forecast horizon of four months. After that period the inclusion of the Google Indicator results in a decrease of the forecasting capability compared to the benchmark model. A visualisation of the performance of both models, considering the forecast performance up till 6 months, is included in Figure A.2 in the Appendix.

4.5 Sub-period analysis

The Google Indicator is improving the forecast accuracy up to four months. However, it is useful to analyze whether that improvement is stable over time or if it varies. Especially the forecasting performances during turning points in the economy are interesting to analyze. According to del Negro (2001) is the ability of models to predict specific events - like turning points - particularly difficult. More recently, a study of Loungani et al. (2013) concludes that the ability of the existing forecast models are not able to predict the start of turning points precisely. According to the previous studies of Choi and Varian (2012) and Tuhkuri (2015) are changes in the search intensity for unemployment related queries able to identify changes in the unemployment rate.

This section analyses whether the inclusion of the Google Indicator can improve the capability of the forecast performance in picking up turning point. This is done by analyzing three turning points in the Dutch Economy. Figure 5 shows the 3 sub-periods considered. According to the CBS are this the periods between January 2004 and April 2017 wherein the most influential changes occurred.

Figure 5: Sub-period forecast analysis



The first period of analysis starts in July 2008. According to the CBS switched the Dutch economy during 2008. The beginning of 2008 was characterized by a stable growth and a decreasing unemployment rate but the recession started in July of that year.[12] This resulted in an increase in the unemployment rate from a historical low point of 3.3% till it reached 5.1% in July 2010.

[12]Source: De Nederlandse Economie 2008. Accessed: 20 May 2017. https://www.cbs.nl/nl-nl/publicatie/2009/37/de-nederlandse-economie-2008

The second period of analysis considers the period between February 2012 until February 2014. This period is characterized by the highest increase in the Unemployment rate in the Netherlands over 30 year. In the beginning of the recession the labour market experienced an increase in the unemployment rate but it stabilized relatively quickly. In the end of 2010 was there even a decrease in the unemployment rate and the labour market seemed to recover. However, it turned out that the consequences of the recession had an delayed effect on the labour market.[13] The unemployment rate rised from February 2012 until February 2014 from 5.6% till 8.4%.

The last period of analysis considers the period between December 2014 and December 2016. Whereas the first two sub-periods assesses an increase in the unemployment rate is this period characterized by a decreasing trend of the unemployment rate. In the beginning of this period the unemployment rate was 7.6%. During the following two years the Netherlands recovered from the recession and the labour market experienced the strongest decline in unemployment rates in over ten years, reaching an unemployment rate of 5.1% in December 2016.[14]

Table 6 shows the result of the analysis of the three sub-periods. This section considers a forecast horizon up to four months to analyze the sub-period performance. As mentioned in the previous section, after four months the Google Augmented model produces less accurate forecast than the benchmark model. This result holds for every sub-period considered.

During the first sub-period the predictive accuracy of the Google augmented model is lower than the benchmark model. Especially the nowcast capability of the Google model is substantially lower with a MAPE of 5.96% compared with a MAPE of 4.04% for the benchmark model. Adding the Google Indicator results in a decrease in the forecasting performance of 47.52%. The predictive accuracy of the Google augmented model stays lower than the benchmark model during every forecast horizon. A possible reason for the decrease in forecasting accuracy of the Google augmented model can be found in the specific time period. As mentioned in section 3.1, the performed Google search queries were substantially lower during the beginning of the period of analysis. This induces an increase in volatile behaviour and this can result in an unstable Google Indicator with a less accurate forecasting capacity.

In comparison with table 5 behaves the benchmark model in a similar way, whereas the Google augmented model experience a substantial increase of the MAPE, especially up to a three month forecast horizon. The differences in forecasting accuracy between both model are not statistically significant according to the D-M test. There are two possible explanations for the insignificant D-M test statistic. First, the observation period is relatively short with 24 forecast observations. Secondly, the differences are relatively small in absolute terms. According to Diebold (2015) has the D-M test lower power in short observation samples, this in combination with the relatively small difference can lead the D-M test to accept the null hypothesis even if the alternative is true.

[13]Source: Nederland in 2014. - een Economisch overzicht. Accessed: 20 May 2017 https://www.cbs.nl/nl-nl/nieuws/2015/13/cbs-presenteert-nederland-in-2014

[14]Source: Nederland in 2016 - een Economisch overzicht. Accessed: 20 May 2017. https://www.cbs.nl/nl-nl/publicatie/2017/12/nederland-in-2016-een-economisch-overzicht

Table 6: Sub-period analysis till 4 months ahead

Sub-period 1	*Model*	*present*	*1 month*	*2 months*	*3 months*	*4 months*
Jul 08 - Jul 10	Benchmark model	4.04%	4.04%	6.71%	8.70%	10.39%
	Google Augmented model	5.96%	4.59%	7.80%	9.89%	10.67%
	Δ	-47.52%	-13.61%	-16.24%	-13.68%	-2.69%
	D-M test statistic	-1.51	-1.03	-0.74	-0.37	-0.07
Sub-period 2						
Feb 12 - Feb 14	Benchmark model	2.67%	2.67%	4.48%	6.22%	7.94%
	Google Augmented model	2.56%	2.51%	4.34%	6.10%	8.11%
	Δ	4.12%	5.99%	3.13%	1.93%	-2.14%
	D-M test statistic	0.26	0.76	0.28	0.28	-0.29
Sub-period 3						
Dec 14 - Dec 16	Benchmark model	3.84%	3.84%	6.05%	7.02%	7.83%
	Google Augmented model	2.51%	3.19%	5.18%	6.19%	7.37%
	Δ	34.64%	16.93%	14.38%	11.82%	5.87%
	D-M test statistic	4.43***	2.71***	2.36**	1.99***	0.77

Forecast is performed using a rolling window of 48 observations.

$\Delta = The\ improvement\ in\ forecasting\ accuracy.$

The second sub-period in table 6 shows the analysis between February 2012 till February 2014. The most substantial changes in the unemployment rate occurred during this period. Interesting to note is that the MAPE of the Google augmented model decreases with approximately 57% in comparison to the first sub-period of analysis, considering the nowcast performance. That improvement indicates that the Google Indicator performs significantly better during the second sub-period of analysis. Considering the MAPE of both models in comparison with table 5, there is an improvement in forecasting accuracy during all forecasting horizons. The marginal improvements of the Google augmented model in comparison to the benchmark are persistent up to a forecast horizon of three-months ahead. However, the magnitude of the improvement in comparison to the benchmark model is modest up to a maximal improvement of 5.99% during the one-month ahead forecast. The relatively small improvements are the reason that the D-M test accepts the null hypothesis of equal forecast accuracy (Harvey et al., 2017).

The last period of analysis considers the recovery of the Dutch labour market and is characterized by a decrease in the unemployment rate from 7.6% till 5.1%. The analysis of the last sub-period in table 6 shows the performance of both models. The improvement of the Google augmented model is higher in comparison to the previous sub-periods. Especially the nowcasting performance of the Google model is substantial higher with a MAPE of 2.51% and a marginal improvement of 34.64% in comparison to the benchmark model. Considering the D-M test, the forecast improvement of the Google augmented model is significant up to the three-months ahead forecast horizon.

Whereas most the existing literature considers one period of analysis explored this section the performance of both models during different time periods. More precisely, the analysis addresses the performance of both models during turning points in the economy. Table 7 shows the evolution of the nowcasting performances of both models. It is clear that the Google augmented model improved substantially over time. The first sub-period analysis uses the observations during June 2004 till June 2008 in order to obtain forecast values for the period July 2008 till July 2010. Adding the Google Indicator decreases the forecast capability substantially. However, the performance of the Google Indicator increases in the following two sub-periods. During the last period of analysis the Google augmented model produces even 34.64% more accurate predictions in comparison with the benchmark model. The evolution of the nowcast performance of both models per month for the three sub-periods is included in figure A.4 in the Appendix.

Table 7: Evolution of nowcasting performance

Time period	*Model*	*MAPE*	Δ
Jul. 2008 - Jul. 2010	Benchmark model	4.04%	
	Google Augmented model	5.96%	-47.52%
Feb. 2012 - Feb. 2014	Benchmark model	2.67%	
	Google Augmented model	2.56%	4.12%
Dec. 2014 - Dec. 2016	Benchmark model	3.84%	
	Google Augmented model	2.51%	34.64%

Forecast is perfomed using a rolling window of 48

$\Delta = The\ improvement\ in\ forecasting\ accuracy.$

The sub-period analysis clearly shows that the usability of a Google Indicator is not stable over time. The results regarding the capacity of the Google augmented models in detecting turning points is not conclusive. The usability of the Google Indicator seems to depend on the time-specific context. A reason for this can be found in the internet penetration in the Netherlands. The use of Google increased between January 2004 and April 2017. As mentioned in the Introduction, the internet penetration in the Netherlands is among the highest in the world. 93% of individuals used the internet on a daily base in 2017. In 2005 this was substantially lower, only 67% of the persons used the internet on a daily base.[15] Besides the increase in internet usage is it as well more often used in order to find a job. According to the CBS used approximately 75% of the unemployed persons the internet as a tool to find another

[15]CBS: Nederland Europees kampioen internettoegang. Accessed: 18 May.
https://www.cbs.nl/nl-nl/nieuws/2012/24/nederland-europees-kampioen-internettoegang

job. In 2004 was that only 40%.[16] During the period of analysis of this study the internet became the main source to obtain information about the labour market. Furthermore, the amount of searches submitted on Google increased substantially. The introduction of the search engine in the Netherlands occured in 2002. in 2012 approximately 50 million search queries were submitted in the Netherlands on a daily base.[17]

The reasons described above indicate that the underlying search volume associated with the unemployment related Google searches increased during the period of analysis. This resulted in a Google indicator based on a index which is more robust towards volatility in the changes in the amount of searches submitted. This increases the forecasting accuracy of the Google Indicator over the years of analysis, which is clearly observable in table 7. This is confirmed by figure A.3 of the Appendix. This figure shows the evolution of the forecast errors and the MAPE of both models during the whole period of analysis.

[16] CBS: Steeds meer werklozen zoeken werk via internet. Accessed: 18 May. urlhttps://www.cbs.nl/nl-nl/nieuws/2012/45/steeds-meer-werklozen-zoeken-werk-via-internet

[17] This number is based on several internet sources. However, Google Sites does not provide the montly search queries nor does it confirm that 50 million is a precise estimate. Therefore the number serves as an insight and need to be interpreted with caution.

5 Robustness checks

In order to validate the performance of the Google Indicator this section presents several robustness checks. First, the strategy of this study in order to obtain the Google Indicator is compared with other strategies. As mentioned in the Literature section, other studies used predefined Google Categories or the single term "Jobs" to construct the Google Indicator. This section analyses if those strategies are producing more accurate forecast models. Thereafter we analyze the sensitivity of the search terms by constructing a Google Indicator based on a search term which is not associated with unemployment benefits or the labour market in general. Finally, we consider rolling windows of several widths.

5.1 Different search term strategies

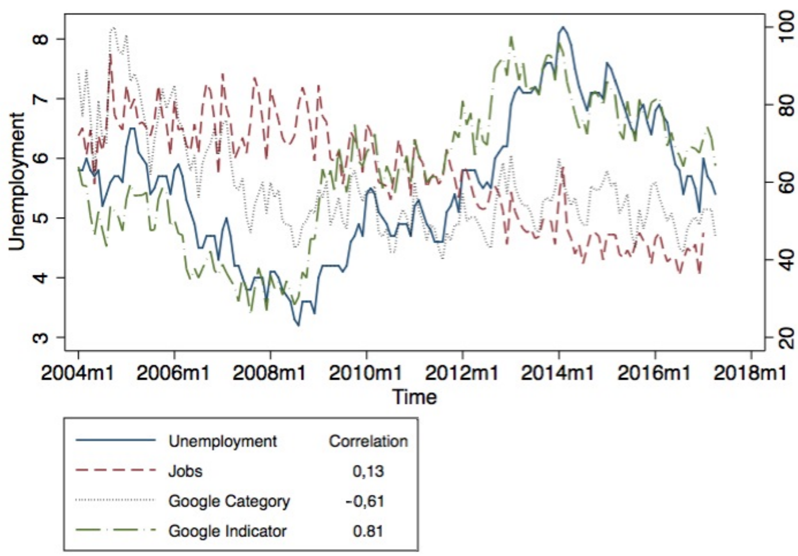
Predefined Google Categories are used among other studies by Choi and Varian (2009a), Suhoy (2009) and Bughin (2011). According to these studies is the underlying search volume the highest by employing the predefined Google categories. Other studies like, D'amuri (2009, 2010, 2012), Fondeur and Karamé (2012) and Vincente et al. (2015) uses solely the keyword "Jobs". These studies motivate their approach by concluding that this particular search terms is the most submitted job-search-related keyword on Google. Additionally, the study of D'amuri (2012) claims that it is widely used across the broadest range of job seekers. Therefore, they suspect that it is less sensitive to sudden changes in searches submitted by a specific subgroup of job seekers.

In order to compare the performance of the augmented model of this study two alternative models are constructed. Instead of the Google Indicator, the baseline model is augmented with a variable associated with the underlying search data of the Google Category: "Local/Jobs". The second alternative model is augmented in a similar way by including the Google Trends data associated with the search terms "Jobs". Figure 6 shows the monthly evolution of the before mentioned models in comparison with the actual unemployment rate and the Google Indicator constructed in this study. It is clearly seen that both of the series of the alternative indicators are not following a similar pattern as the actual unemployment rate. This is confirmed by the correlation coefficient. The correlation between the Unemployment rate and the searches submitted for "Jobs" is 0.13. There is a negative correlation between the Google categories and the unemployment rate of 0.61. This indicate that the series evolve in an opposite way.

The result of the in-sample analysis are included in the Appendix in table A.4. The in-sample analysis shows that the weighted Google Indicator has a higher magnitude in comparison with the other variables. Furthermore, the information criteria of the model augmented with the Google Indicator is smaller than the other models. This result confirms that the model augmented with the Google Indicator of this study is the parsimonious model.

To analyze whether the forecast produced by the model augmented with the Google Indicator outperforms the other models, regressions are performed. For reasons of brevity are the forecast performance of the model considering an extended forecast horizon included in the Appendix in table A.5. The nowcast performance of the different models is shown in table 8. The forecast are produced using a rolling window of 48 months in order to evaluate the forecast period between

Figure 6: Monthly evolution of the unemployment rate and the different Google Indicators



January 2015 - January 2017. As expected, all three of the augmented models are producing more accurate nowcast than the benchmark model. The measurement tools are showing the lowest numbers for the model augmented with the Google Indicator, this implies that the model with the Google Indicator produces the most accurate forecasts. However, considering this period of analysis, the model augmented with the search term "Jobs" is producing nowcast which have similar accuracy power than the model with the Google Indicator. According to the D-M test statistic are both models significantly more accurate than the benchmark model and the model augmented with the predefined Google Categories. The accuracy power of the model augmented with the search term "Jobs" is decreasing after a forecast period of one month. This is not the case for the model augmented with the Google Indicator. The Google augmented model outperforms the alternative models during all forecast horizons considered.

Table 8: A comparison of the One-step ahead forecast performance

Model	*MAE*	*RMSE*	*MAPE*	Δ
Benchmark model	0.280	0.331	4.53%	
Google Augmented model	0.200	0.255	3.22%	28.92%
Google Category model	0.238	0.284	3.85%	15.01%
Google "Jobs" model	0.201	0.260	3.23%	28.70%

$\Delta = The\ improvement\ in\ forecasting\ accuracy.\ versus\ the\ benchmark\ model$

The evaluation period is January 2015 - January 2017. Forecast is perfomed using a rolling window of 48 observations, started from January 2013.

5.2 A fake Google Index

A possible concern is the possibility that the results are obtained solely by including an additional variable. Therefore the sensitivity of the obtained results are analyzed by constructing a variable based on the search index of a term which is not related with unemployment or the labour market in general.

The fake google Index is based on the most submitted search term during the period of analysis in the Netherlands; "YouTube".[18] An unrelated search term should not have any explanatory power. The results of the in-sample analysis are included in the Appendix in table A.6. The regression shows that the variable is not significant. This confirms that only Google search terms related to the labour market have explanatory power and are able to improve the forecasting models.

5.3 Alternative window width

Throughout this study the most common rolling forecast window of 48 periods is mainly used. This approach is in line with, among others, the studies of Choi and Varian (2012), Pavlicek and Kristoufek (2015) and Tuhkuri (2015). The window size of 48 periods ensures that enough observations are used in order to estimate the models, and it makes that the forecast evaluation period is long enough.

To analyze the sensitivity of the forecast performance of both models we selected window sizes of 24, 36 and 60 months. The result of the performed regressions are included in table A.7, A.8 and A.9 of the Appendix. The results show that the Google augmented model is robust against different window widths. The period of analysis is January 2015 till April 2017. The Google augmented models tend to even further improve in performance when the window width is expanded. This confirms the previous results. As mentioned in the sub-period analysis section, the Google augmented model is becoming more accurate over time. The period of analysis uses the most recent observations and by including more (recent) observations the forecast performance increases. On the other hand, the performance of the benchmark model is not improving but stay rather stable when the window size is expanded.

A final robustness check considers the specification of the benchmark model. In line with most of the previous literature an autoregressive model of order 1 is used as the benchmark. This specification is typically used to model the unemployment rate as pointed out by Montgomery (1998). The alternative specification of the benchmark model includes an aditional variable, the seasonal autoregessive term. As the result in table A10 of the Appendix shows, the Google augmented model outperforms as well the benchmark model which include the seasonal term. This result support the relatively simple benchmark model employed in this study.

[18]The most submitted search term is found by exploring the top 10 of the Google Trends search words for the years between 2004-2016. Accessed: 30 May.
https://trends.google.com/trends/topcharts#geo=NL&date=2015

6 Conclusions

The last years have witnessed how the use of internet has become a major element in daily life. In the Netherlands 82.1% of the population uses the internet on a daily base and it has become the major source for job-seekers to find a job. Since 2007 Google Sites makes the search behaviour publicly available by introducing the Google Trends Application. This study analysed the usability of the Google Trends data in order to predict the unemployment in the Netherlands. Therefore a benchmark model and a Google augmented model are constructed and the performances are compared.

The main result of this study indicates that using the google queries improves the forecast accuracy. The first result of this study uses the data from January 2004 till December 2014 in order to obtain one-month ahead forecast of the unemployment rate during the period between January 2015 till April 2017. In order to compare the predictive capability of both models the MAPE is used. The outcome shows an improvement of the forecast accuracy of 8.7%. This result is in line with the studies of Choi and Varian (2012) and Tuhkuri (2015), however they analyzed the labour market of the USA.

When considering a rolling forecast with a window of 48 observations we find an even higher increase of the performance of the Google augmented model in comparison to the benchmark model. During the period between January 2015 and April 2017 we obtain a MAPE of 3.27% for the Google augmented model, compared with a MAPE of 4.53% for the benchmark model. This results shows a marginal improvement of 27.81%, considering a one-month ahead forecast. The magnitude of this result is in line with the research of D'amuri (2012) and Vicente (2015) who reported results for the USA and Spain respectively.

Most of the previous research is focused on the one-month ahead forecast capacity of the models. This study analyses as well the performance of the Google Indicator during a longer forecast period. The results shows that the Google augmented model outperforms the benchmark model up to a forecast horizon of three months. This indicates that the Google augmented model not only produces more accurate nowcast predictions but outperforms the benchmark model as well in the short-term.

A more in-dept analysis of the performance of the Google augmented model shows that the before mentioned results needs to be interpreted with caution. The sub-period analysis shows that the performance of the Google augmented model is not stable over time. During the first sub-period, which considers July 2008 till July 2010, the Google augmented model produces less accurate forecast in comparison to the benchmark model, with a MAPE of 5.96% versus a MAPE of 4.04%. Over time we observe a substantial improvement in the performance of the Google augmented model. The last sub-period analyzes the performance of both models during December 2014 till December 2016. The results show that the MAPE of the Google augmented model decreased to 2.51% while the performance of the benchmark model remained relatively stable over time. This means that the Google augmented produces during this period forecast which are 34.64% more accurate than the benchmark model. Despite the relatively low amount of sub-periods we observed a trend of an improving performance of the Google augmented model. We suspect that the underlying reason for this improvement can be found in the increase of internet usage, more precisely Google usage, during the period of analysis. A higher amount of submitted searches results in a more stable Google Indicator which is capable to produce more accurate forecast.

A final result of this study can be found in the comparison of different Google augmented models. This is the first study which analyses different types of Google augmented models. Besides the Google Indicator obtained in this study, regressions are performed based on predefined Google categories and the single search term "Jobs". The results shows that in the case of the Netherlands the Google Indicator of this study, which is based on multiple search terms related to unemployment benefits, produces the most accurate forecast.

Overall, the presented evidence highlights the importance and usefulness of Google Trends data for the nowcasting and forecasting models. The performance of the Google augmented models are becoming even more accurate during the recent years. This motivates to include the Google Trends data in the existing forecast models, especially because the internet plays a major role in the economy (Einav and Levin, 2013) and we suspect that the increase and accessibility of internet-based data will expand even more in the near future.

Three final remarks regarding the results of this study should be made. First, despite the convincing advantages of using internet-based, more precisely Google Trends, data is their a risk of capturing a certain amount of noise in the data. In other words, we can not be completely sure that the underlying reason for each unemployment related search term submitted is because of the possibility of becoming unemployed. More sophisticated variable selection methods like the Bayesian Variable Selection could be used to reduce the possible amount of noise. That method is designed specifically to analyze search engine data.

Secondly, although the underlying data of this studies includes more observation then previous research, the observation period remains relatively short with 160 observations in total. The observed trend of an improving performance of the Google augmented model is based on the last 96 observations, it is not clear whether this pattern will continue in the future.

The final remark of this studies involves the external validity of the result. The procedure put forward in this research is difficult to replicate for other countries. The results depends on native words, a specific time period and composition and search behavior of the Google users in the Netherlands. Nevertheless, all the previous research concludes that Google augmented models tend to outperform the models which exclude the Google variable. The magnitude of the improvement varies per study and country and depends as well on the modeling choice and instruments employed.

References

Anvik, C. and Gjelstad, K. "just google it": Forecasting norwegian unemployment figures with web queries. *Working Paper CREAM - BI Norwegian Business School*, 2010.

Armstrong, J. *Principles of Forecasting.* Kluwer Academic publishers, 2002.

Aruoba, S. and Diebold, F. Real-time macroeconomic monitoring: Real activity, inflation, and interactions. *American Economic Review*, 100(2):20 – 24, 2010.

Askitas, N. and Zimmerman, K. Google econometrics and unemployment forecasting. *Applied Economics Quaterly*, 55(2):107–120, 2009.

Bijl, L., Kringhaug, G., Molnár, P., and Sandvik, E. Google searches and stock returns. *International Review of Financial Advice*, 45(3):150–156, 2016.

Box, G., Jenkins, G., Reinsel, G., and Ljung, G. *Time Series Analysis: Forecasting and Control.* Wiley, 5th edition, Augustus 2015.

Breur, T. Big data and the internet of things. *Journal of Marketing Analytics*, 3(1):1 – 4, 2015.

Brodersen, K., Gallusser, F., Koehler, J., Remy, N., and Scott, S. Inferring causal impact using bayesian structural time-series models. *Annals of Applied Statistics*, 9:247 – 274, 2015.

Bughin, J. "nowcasting" the belgian economy. *SSRN Electronic Journal*, 2011.

Chadwick, M. G. and Sengül, C. G. Nowcasting unemployment rate in turkey: Let's ask google. *Research and Monetary Policy Department - Central Bank of the Republic of Turkey*, 2012.

Chai, T. and Draxler, R. Root mean square error (rmse) or mean absolute error (mae)? – arguments against avoiding rmse in the literature. *Geoscientific Model Development Discussions*, 7(1):1525 – 1534, 2014.

Choi, H. and Varian, H. Predicting initial claims for unemployment benefits. *Technical report Google*, 2009a.

Choi, H. and Varian, H. R. Predicting the present with google trends. *Economic Record*, 88(1): 2–9, 2012.

Choi, H. and Varian, H. Predicting the present with google trends. *Technical report Google*, 2009b.

D'Amuri, F. Predicting unemployment in short samples with internet job search query data. *Bank of Italy - Research Department*, 2009.

D'Amuri, F. and Marcucci, J. "google it!" forecasting the us unemployment rate with a google job search index. *Bank of Italy Working Paper*, 2010.

D'Amuri, F. and Marcucci, J. The predictive power of google searches in forecasting unemployment. *Bank of Italy Working Paper 891*, 2012.

Diebold, F. Comparing predictive accuracy, twenty years later: A personal perspective on the use and abuse of diebold–mariano tests. *Journal of Business Economic Statistics*, 33(1):1 – 10, 2015.

Dugas, A., Jalalpour, M., Gel, Y., Levin, S., Torasco, F., Igusa, T., and Rothman, R. Influenza forecasting with google flu trends. *PLoS ONE*, 8(2), 2013.

Einav, L. and Levin, J. The data revolution and economic analysis. *NBER Working Paper 19035*, 2013.

Ettredge, M., Gerdes, J., and Karuga, G. Using web-based search data to predict macroeconomic statistics. *Communications of the ACM*, 48(11):87–92, 2005.

Faberman, R. and Kudlyak, M. What does online job search tell us about the labor market? *Economic perspectives*, 40(1), 2016.

Falorsi, S., Naccarato, A., and Pierini, A. Using google trend data to predict the italian unemployment rate. *Working Paper - Dipartimento di Economia Università degli studi Roma Tre*, 203, 2015.

Franses, P. A note on the mean absolute scaled error. *International Journal of Forecasting*, 32: 20 – 22, 2016.

Gunter, U. and Önder, I. Forecasting city arrivals with google analytics. *Annals of Tourism Research*, 61:199–212, 2016.

Hamid, A. and Heiden, M. Forecasting volatility with empirical similarity and google trends. *Journal of Economic Behavior Organization*, 117:62–81, 2015.

Harvey, D., Leybourne, S., and Whitehouse, E. Forecast evaluation tests and negative long-run variance estimates in small samples. *International Journal of Forecasting*, In Press, 2017.

Hyndman, R. and Koehler, A. Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22:679 – 688, 2006.

Koop, G. and Onorante, L. Macroeconomic nowcasting using google probabilities. *Working paper - European Central Bank*, 2013.

Kuhn, P. and Mansour, H. Is internet job search still ineffective? *The Economic Journal*, 124: 1213–1233, 2014.

Lasso, F. and Snijders, S. The power of google search data; an alternative approach to the measurement of unemployment in brazil. november 2016.

Mauro, A. D., Greco, M., and Grimaldi, M. A formal definition of big data based on its essential features. *Library Review*, 65(3):122–135, 2016.

McLaren, N. and Shanbhogue, R. Using internet search data as economic indicators. *Bank of England Quaterly Bulletin*, 51(2):134–140, 2011.

Montgomery, A., Zarnowitz, V., Tsay, R., and Tiao, G. Forecasting the u.s. unemployment rate. *Journal of Americal Statistical Association*, 93, 1998.

Negro del, M. Turn, turn, turn: Predicting turning points in economic activity. *Federal Reserve Bank of Atlanta - Economic Review - Working Paper*, 2001.

Newey, W. and West, K. Automatic lag selection in covariance matrix estimation. *Review of Economic Studies*, 61:631 – 653, 1994.

Nyberg, H. Dynamic probit models and financial variables in recession forecasting. *Journal of Forecasting*, 29:215 – 230, 2010.

O'Leary, D. E. 'big data', the 'internet of things' and the 'internet of signs'. *Intelligent Systems in Accounting, Finance and Management*, 20(50 - 65), 2013.

Pavlicek, J. and Kristoufek, L. Nowcasting unemployment rates with google searches: Evidence from the visegrad group countries. *PLoS ONE*, 10(5), 2015.

Preis, T., Maat, H., and Stanley, H. Quantifying trading behavior in financial markets using google trends. *Scientific Reports*, 3(1684):1 – 6, 2013.

Rivera, R. A dynamic linear model to forecast hotel registrations in puerto rico using google trends data. *Tourism Management*, 57:12 – 20, 2016.

Scott, S. and Varian, H. Bayesian variable selection for nowcasting economic time series. *University of Chicago Press*, National Bureau of Economic Research:118 – 135, 2015.

Su, Z. Chinese online unemployment-related searches and macroeconomic indicators. *Frontiers of Economics in China*, 9(4):573 – 605, 2014.

Suhoy, T. Query indices and a 2008 downturn: Israeli data. *Research Department - Bank of Isreal*, 2009.

Teng, Y., Bi, D., Xie, G., Jin, Y., Huang, Y., Lin, B., An, X., Feng, D., and tong, Y. Dynamic forecasting of zika epidemics using google trends. *PLoS ONE*, 2017.

Tuhkuri, J. Big data: Do google searches predict unemployment? Master's thesis, University of Helsinki, 2015.

Varian, H. R. Computer mediated transactions. *The American Economic Review*, 100(2):1–10, 2010.

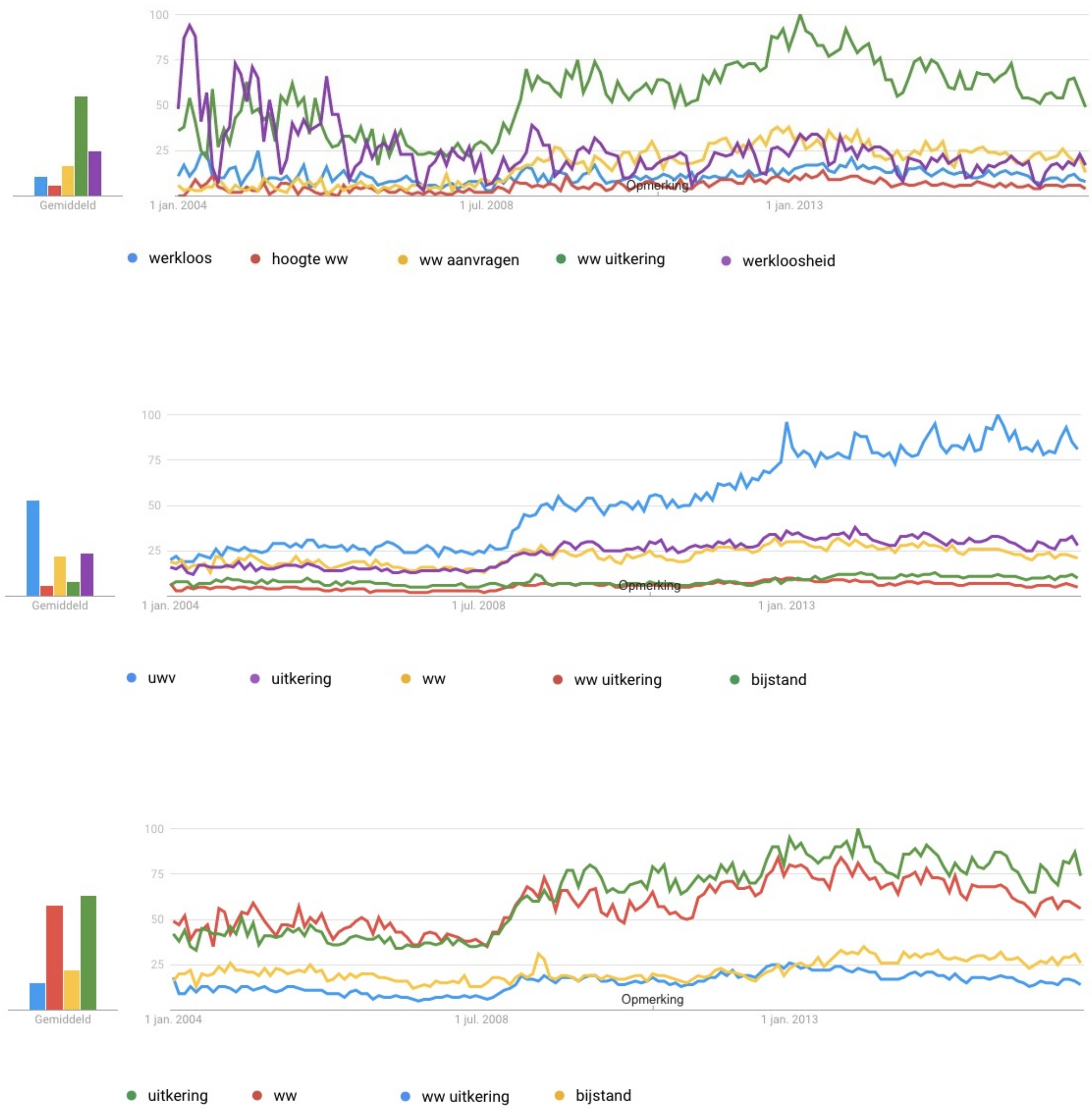
Veldhuizen, van, S., Vogt, B., and Voogt, B. Internet searches and transactions on the dutch housing market. *Applied Economics Letters*, 23(18):1321 –1324, 2016.

Vicente, M., López-Menéndez, A., and Pérez, R. Forecasting unemployment with internet search data: Doet it help to improve predictions when job destruction is skyrocketing? *Technological Forecasting Social Change*, 92:132–139, 2015.

Willmott, C. and Matsuura, K. Advantages of the mean absolute error (mae) over the root mean square error (rmse) in assessing average model performance. *Climate research*, 30(1):79 – 82, 2005.

7 Appendix

Figure A1: Boolean Operator



The figure above shows the search volume associated with the search terms. Most of the searches are submitted for the word “UWV”, therefore this word has the highest weight. The last graphic shows the change in the Index when the search term “UWV” is excluded. The weight used in this study are based on different combinations of the Boolean Operator and the search volume range provided by Google AdWords.

Table A1: Overview of previous work - part 1

Empirical studies analyzing the unemployment rate[1]					
No.	*Author (publ.)*	*Country*	*Google data usage*	*Time horizon*	*Improved forecast accuracy*
1	Choi (2009)	USA	Google categories	1 month	MAE decrease of 3.3%
2*	Askitas (2009)	Germany	Multiple search terms	1, 2, 3 months	n.a. [2]
3	Suhoy (2009)	Israel	Google categories	1 month	RMSE decrease of 27.2%
4*	D'amuri (2009)[3]	Italy	Single word	1, 2, 3 months (quaterly data)[4]	1 month: MSE decrease of 26.9% 2 months: MSE decrease of 50%. 3 months: MSE decrease of 42.4%
5	Anvik (2010)	Norway	Multiple search terms	1 month	RMSE decrease up to 15.3%
6	D'amuri (2010)	USA	Single word	1 month	RMSE decrease of 40%
7	M'Claren (2011)	England	Single word	1 month	RMSE decrease of 12.6%
8	Bughin (2011)	Belgium	Google categories	1 month	Google Search behavior explains 15% of unemployment movements. No forecast comparison is made with a benchmark model.
9	Fondeur (2012)	France	Single word	1 month	RMSE decrease of nearly 27%.
10*	Chadwick (2012)	Turkey	Used Bayesian Model Average(BMA) to obtain several combinations of multiple search terms.	1 quarter	MSE decrease of 38.3%

* *Denotes working papers*

[1] The table only considers studies where the variable of interest is the unemployment rate and the data is obtained by using the Google Trends application.

[2] The authors did not quantify a comparison of the forecasting accuracy. Their objective consist out of finding a correlation.

[3] The studies of D'amuri uses horse-races between 300-500 different models. This table compares the best Google augmented model with the best model which exclude the Google Indicator. For all the different outcome see his papers.

[4] This indicates one month nowcasting extended with two monts of forecasting.

Table A.1 :Overview of previous works - part 2

No.	*Author (publ.)*	*Country*	*Google data usage*	*Time horizon*	*Improved forecast accuracy*
11	Choi (2012)	USA	Google categories	1 month	MAE increase of 5.95% MAE decreses with 13.6% during the recession.
12	D'amuri (2012)	USA	Single word	1 month	MSE decrease of 28%
13	Su (2014)	China	Single word	1 month	n.a. [5]
14	Vicente (2015)	Spain	Single word	1 month	RMSE decrease of 26.8%
15	Pavlicek (2015)	Visegrad Group countries	Single word for each country	1 month	The D-M statistics shows that Google augmented model is significantly more accurate for; Czech. Rep, Hungary and Poland. This is not the case for Slovakia month
16*	Falorsi (2015)	Italy	Single word	1 month	MSE decrease of 83.3%
17	Tuhkuri (2015)[6]	USA	Multiple search terms	7 months	1 month: MAPE decrease of 4.32% 2 months: MAPE decrease of 7.48% 3 months: MAPE increase of 3.92% 4 months: MAPE increase of 6.28% 5 months: MAPE increase of 17.22% 6 months: MAPE increase of 13.22% 7 months: MAPE decrease of 9.93%
18	Lasso (2016)	Brazil	Multiple search terms	1 month	n.a.[7]

* *Denotes working papers*

[5] The assesment of this studies is the analysis of a correlation between unemployment series and Google search behavior. The results shows significant correlations.

[6] This studies analysis as well turning points in the economy. The accuracy of the Google augmented models is approximately four times larger during those moments than the results reported in the table.

[7] The authors did not quantify a comparison of the forecasting accuracy. Their objective consist out of finding a correlation.

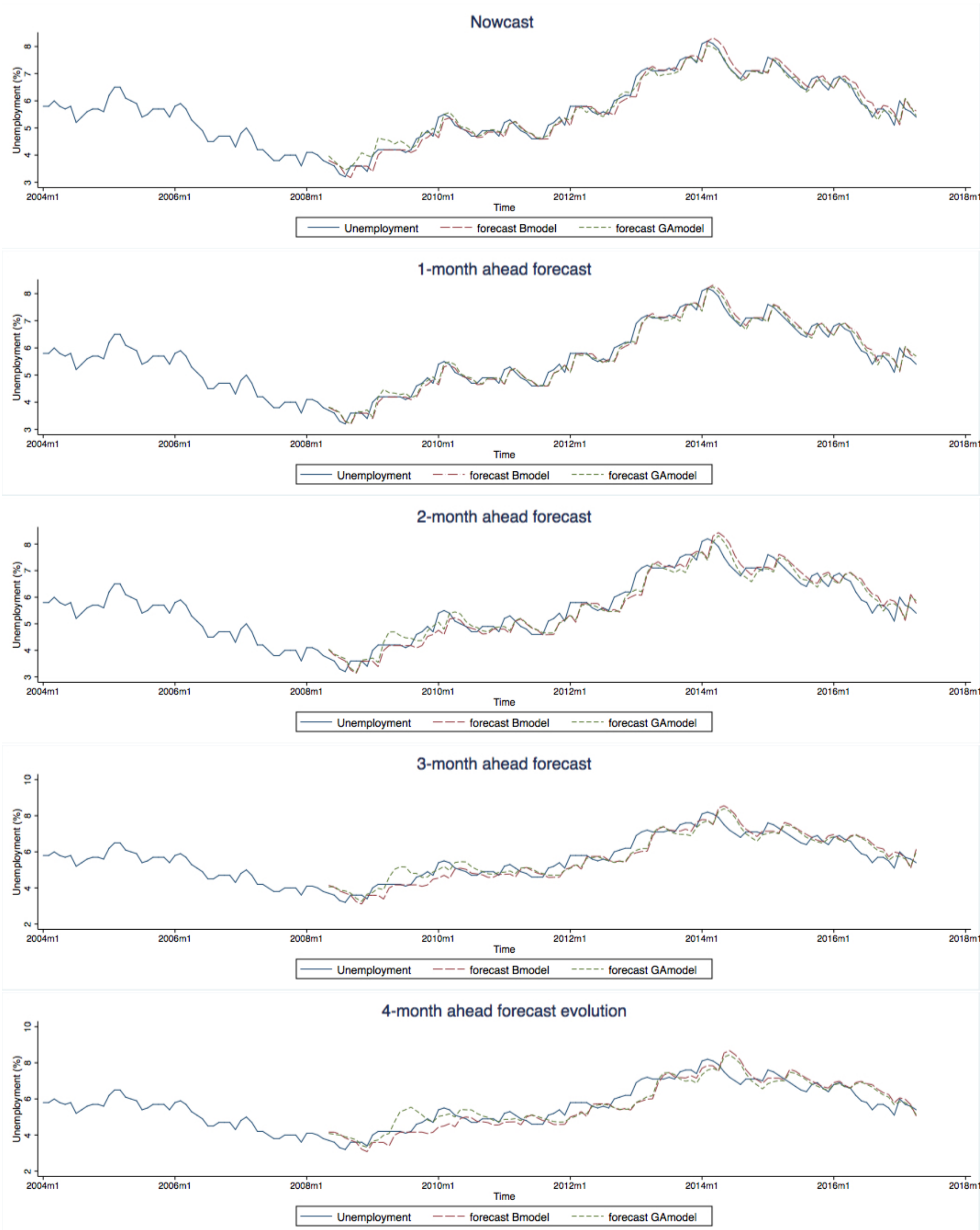
Table A.2: Comparison of the forecast accuracy over time - one month ahead

Time	*Actual Data*	*Forecast Bmodel*	*Forecast GAmodel*	*MAE Bmodel*	*MAE GAmodel*	*MAPE Bmodel*	*MAPE GAmodel*
01/2015	7.6	7.01	7.08	0.59	0.52	7.72	6.90
01/2015	7.6	6.98	7.05	0.62	0.55	8.14	7.24
02/2015	7.5	7.59	7.57	0.09	7.00	1.17	0.94
03/2015	7.3	7.49	7.46	0.19	0.16	2.55	2.25
04/2015	7.1	7.28	7.19	0.18	0.09	2.55	1.3
05/2015	6.9	7.08	7	0.18	0.1	2.64	1.4
06/2015	6.7	6.88	6.91	0.18	0.21	2.73	3.19
07/2015	6.5	6.68	6.68	0.18	0.18	2.81	2.72
08/2015	6.4	6.49	6.46	0.09	0.06	1.4	0.98
09/2015	6.8	6.39	6.48	0.41	0.32	6.04	4.75
10/2015	6.9	6.79	6.79	0.11	0.11	1.66	1.56
11/2015	6.6	6.89	6.86	0.29	0.26	4.32	4.01
12/2015	6.4	6.59	6.64	0.19	0.24	2.9	3.79
01/2016	6.8	6.39	6.47	0.41	0.33	6.06	4.81
02/2016	6.9	6.78	6.83	0.12	7.00	1.73	1.10
03/2016	6.7	6.88	6.85	0.18	0.15	2.65	2.30
04/2016	6.6	6.68	6.6	0.08	0	1.18	0.07
05/2016	6.2	6.58	6.55	0.38	0.35	6.19	5.66
06/2016	5.9	6.19	6.16	0.29	0.26	4.88	4.42
07/2016	5.8	5.89	5.88	0.09	0.08	1.59	1.44
08/2016	5.4	5.8	5.77	0.4	0.37	7.36	6.84
09/2016	5.7	5.4	5.48	0.3	0.22	5.24	3.93
10/2016	5.7	5.7	5.73	0	0.03	0.07	0.59
11/2016	5.5	5.7	5.73	0.2	0.23	3.57	4.16
12/2016	5.10	5.5	5.54	0.4	0.44	7.82	8.71
01/2017	6	5.11	5.24	0.89	0.76	14.9	12.6
02/2017	5.7	5.99	6.04	0.29	0.34	5.10	6.02
03/2017	5.6	5.69	5.75	0.09	0.15	1.68	2.76
04/2017	5.4	5.6	5.58	0.2	0.18	3.65	3.39

The evaluation period is January 2015 - April 2017. Forecast is performed using a rolling window of 132 observations, started from January 2004.

Bmodel = Benchmark model. GAmodel= Google Augmented model.

Figure A.2: Evolution of the forecast accuracy of the benchmark model and the Google augmented model in comparsion to the actual unemployment rate[8]



[8]The forecast is performed using a rolling window of 48 observations.

Figure A.3: Evolution of the Forecast errors and the MAPE of the benchmark model and the Google augmented model

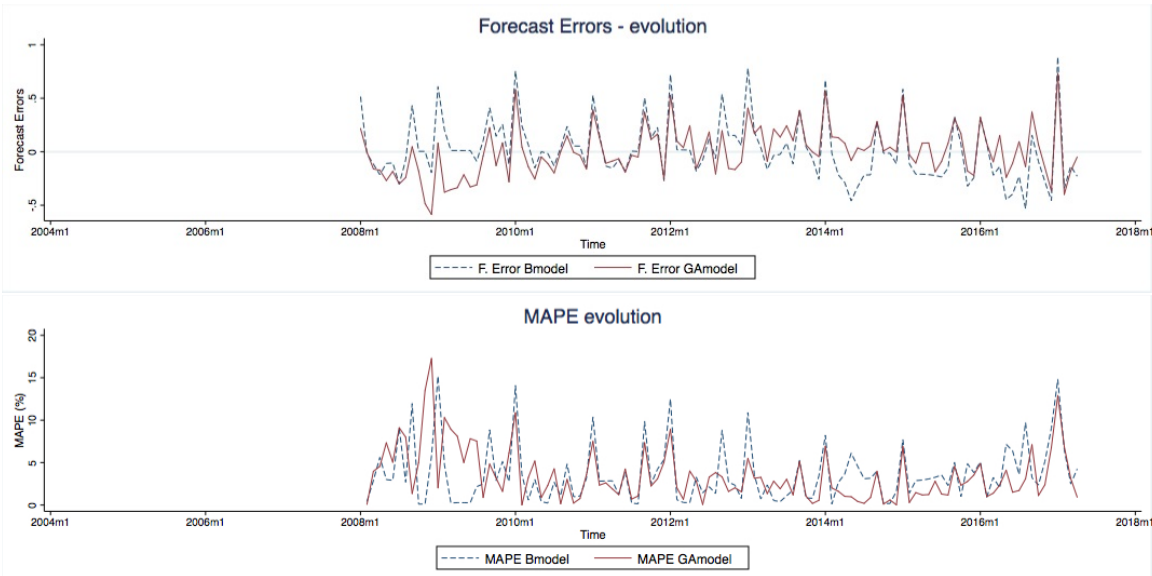
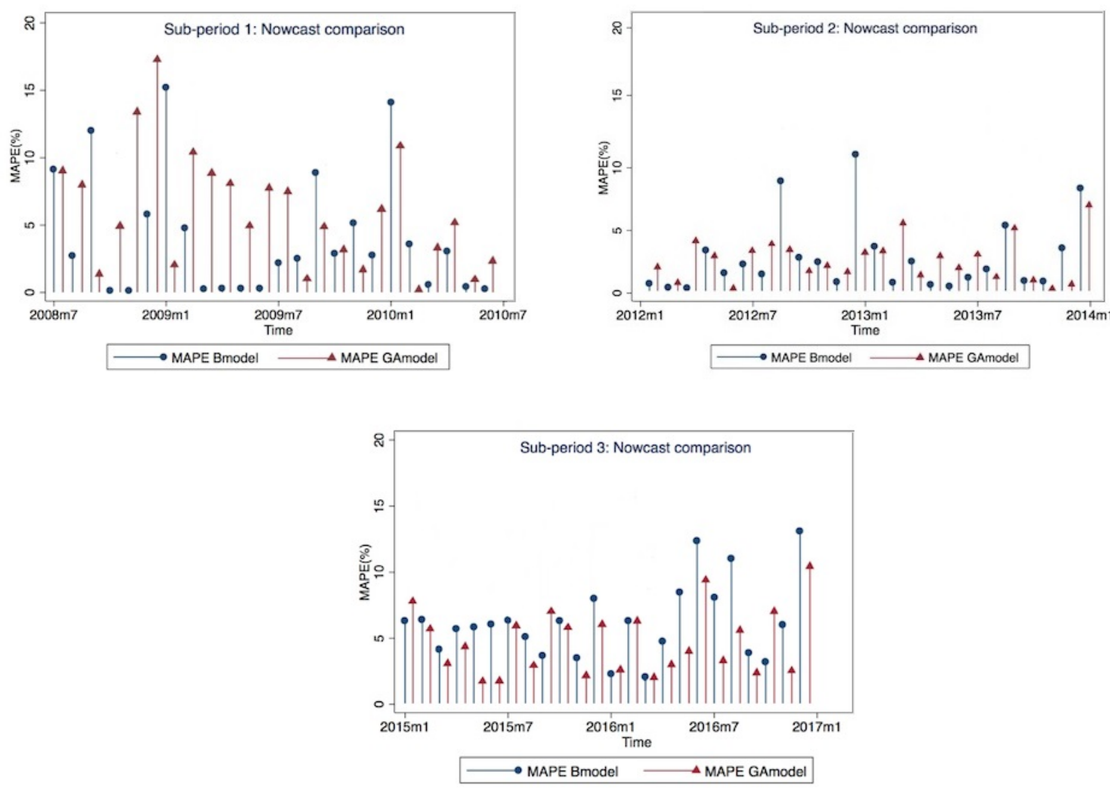


Figure A.4: Evolution of the MAPE during the sub-period analysis



The forecasts in Figure A.3 and Figure A.4 are performed using a rolling window of 48 observations.

Table A4. Estimation results of the different Google Indexes

Variable	*Benchmark model*	*Google Indicator*	*Google Categories*	*"Jobs"*	*All Var.*
U_{t-1}	0.974^{***}	0.857^{***}	0.969^{***}	1.002^{***}	-1.030^{***}
GI		0.010^{***}			0.155^{***}
GC.			0.004^{**}		0.005^{**}
Jobs.				0.005^{***}	0.008^{***}
Cons.	0.144	0.163^{*}	-0.075	-0.510^{**}	-1.030^{***}
Summary					
$\bar{R}^2$	0.948	0.959	0.950	0.951	0.971
BIC	41.1	9.3	30.6	27.2	-38.4
AIC	34.9	0.1	39.8	36.4	-53.8
N	159	159	159	159	159

Stardard errors in the parenthesis $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

U = Unemployment, GI = Google Indicator GC = Google Categories

Table A5: Comparison of the MAPE performance of the different types of Google augmented models

Model	*present*	*1 month*	*2 months*	*3 months*	*4 months*
Benchmark model	4.53%	4.53%	6.87%	7.55%	7.66%
Google Augmented model	3.27%	4.09%	6.07%	6.66%	6.98%
Google Category model	3.85%	4.55%	6.87%	7.61%	7.76%
Google "Jobs" model	3.23%	4.61%	6.90%	7.32%	7.53%

Forecast is performed using a rolling window of 48 observations, - April 2017.

Table A6. Estimation results of the Fake Google Index

Variable	*Benchmark model*	Fake Index"Youtube"
U_{t-1}	0.974^{***}	0.857^{***}
YT		0.001
Cons.	0.144	0.148
Summary		
$\bar{R}^2$	0.948	0.949
BIC	41.1	43.51
AIC	34.9	34.3
N	159	159

Stardard errors in the parenthesis $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

U = Unemployment, YT = Youtube

Table A.7: MAPE performance during different forecast horizons

Model	*present*	*1 month*	*2 months*	*3 months*	*4 months*	*5 months*	*6 months*
Benchmark model	4.57%	4.56%	6.86%	7.37%	6.82%	8.76%	10.1%
Google Augmented model	3.26%	4.54%	6.63%	6.80%	6.92%	13.43%	13.1%
Δ	28.67%	0.44%	3.35%	7.73%	-1.47%	-64.73%	-29.70%
D-M test statistic	4.28***	0.05	0.96	2.26**	-0.58	-2.32**	-3.44**

Forecast is performed using a rolling window of 24 observations, the evaluation period is January 2015 - April 2017.

$\Delta = The\ improvement\ in\ forecasting\ accuracy.$ ***p<0.01**p<0.05, *p<0.1.

Table A.8: MAPE performance during different forecast horizons

Model	*present*	*1 month*	*2 months*	*3 months*	*4 months*	*5 months*	*6 months*
Benchmark model	4.44%	4.44%	6.69%	7.47%	7.04%	9.33%	10.56%
Google Augmented model	3.24%	4.30%	6.31%	6.50%	7.18%	11.39%	12.27%
Δ	27.03%	3.16%	5.69%	12.99%	-1.99%	-22.08%	-16.20%
D-M test statistic	3.70***	0.58	1.30	2.35**	-0.72	-2.81**	-3.09***

Forecast is performed using a rolling window of 36 observations, the evaluation period is January 2015 - April 2017.

$\Delta = The\ improvement\ in\ forecasting\ accuracy.$ ***p<0.01**p<0.05, *p<0.1.

Table A.9: MAPE performance during different forecast horizons

Model	*present*	*1 month*	*2 months*	*3 months*	*4 months*	*5 months*	*6 months*
Benchmark model	4.45%	4.45%	6.62%	7.17%	7.04%	8.77%	10%
Google Augmented model	3.07%	3.89%	5.79%	6.50%	6.73%	9.63%	10.49%
Δ	31.01%	12.58%	12.54%	9.34%	4.41%	-9.81%	-4.90%
D-M test statistic	4.92***	3.06***	2.60***	1.93*	0.71	-1.81*	-1.89*

Forecast is performed using a rolling window of 60 observations, the evaluation period is January 2015 - April 2017.

$\Delta = The\ improvement\ in\ forecasting\ accuracy.$ ***p<0.01**p<0.05, *p<0.1.

Table A.10: MAPE performance of the Seasonal adjusted models during different forecast horizons

Model	*present*	*1 month*	*2 months*	*3 months*	*4 months*	*5 months*	*6 months*
Benchmark model	4.56%	4.56%	6.86%	7.37%	6.82%	8.76%	10.1%
Benchmark model Seasonal	4.57%	4.30%	7.06%	7.60%	8.03%	10.24%	11.66%
Google Augmented model	3.26%	4.54%	6.63%	6.80%	6.92%	13.43%	13.1%
Google Augmented model Seasonal	3.38%	4.32%	6.11%	6.52%	7.41%	11.42%	12.82%

Forecast is performed using a rolling window of 48 observations, the evaluation period is January 2015 - April 2017.