International Journal of Forecasting 33 (2017) 801–816

Contents lists available at ScienceDirect

International Journal of Forecasting

journal homepage: www.elsevier.com/locate/ijforecast

The predictive power of Google searches in forecasting US unemployment

Francesco D'Amuri [a,b], Juri Marcucci [a,*]

[a] Bank of Italy, Italy
[b] IZA, Italy

ARTICLE INFO

Keywords:
Google econometrics
Forecast comparison
Keyword search
US unemployment
Time series models

ABSTRACT

We assess the performance of an index of Google job-search intensity as a leading indicator for predicting the monthly US unemployment rate. We carry out a deep out-of-sample forecasting comparison of models that adopt the Google Index, the more standard initial claims, or alternative indicators based on economic policy uncertainty and consumers' and employers' surveys. The Google-based models outperform most of the others, with their relative performances improving with the forecast horizon. Only models that use employers' expectations on a longer sample do better at short horizons. Furthermore, quarterly predictions constructed using Google-based models provide forecasts that are more accurate than those from the Survey of Professional Forecasters, models based on labor force flows, or standard nonlinear models. Google-based models seem to predict particularly well at the turning point that takes place at the beginning of the Great Recession, while their relative predictive abilities stabilize afterwards.

1. Introduction

The provision of accurate predictions of labor market dynamics has always been a core activity for both investors and policy makers. The task has become even more important since the beginning of the Great Recession and the related uncertainty regarding the impact of the slowdown in economic activity on the labor market. It then became crucial in December 2012, when the Federal Reserve announced a shift in the way in which monetary policy is communicated to the public. The Fed moved from indicating monetary actions based on time, to explicit employment and inflation guideposts (the so-called Evans rule). For the employment guidepost, the Fed explicitly formulated an unemployment rate threshold of 6.5%, above which a federal funds rate hike is unlikely.

Against this background, we assess whether US monthly unemployment rate predictions can be improved using the Google index (GI), a leading indicator that is based on internet job-related searches performed through Google.[1]

* Correspondence to: Bank of Italy, Directorate General for Economics, Statistics and Research, Via Nazionale 91, 00184, Rome, Italy.

E-mail addresses: francesco.damuri@bancaditalia.it (F. D'Amuri), juri.marcucci@bancaditalia.it (J. Marcucci).

[1] The US unemployment rate time series is certainly one of the most commonly studied series in the literature. Proietti (2003) defines this series as the 'testbed' or 'case study' for many (if not most) nonlinear time series models. In fact, many papers have documented its asymmetric behavior. DeLong and Summers (1986), Neftci (1984) and Rothman (1998) document a type of asymmetry called *steepness*, in which unemployment rates rise faster than they decrease. Sichel (1993) finds evidence of another type of asymmetry called *deepness*, in which contractions are deeper than expansions. McQueen and Thorley (1993) find *sharpness*, in which peaks tend to be sharp while troughs are usually more rounded. In a recent paper, Barnichon and Nekarda (2012) develop a model based on labor force flows for forecasting unemployment; their results indicate that this approach can improve the forecast accuracy of standard time series models considerably.

http://dx.doi.org/10.1016/j.ijforecast.2017.03.004
0169-2070/

After having selected the best specifications at each forecast origin using the BIC, we test the predictive power of this indicator by means of a deep out-of-sample comparison carried out along two different dimensions: (i) the alternative exogenous variables adopted as leading indicators; and (ii) the length of the estimation sample. In particular, we estimate standard time series (AR) models and augment them with different leading indicators such as the initial claims, employers' and consumers' surveys on employment dynamics, the economic policy uncertainty index of Baker, Bloom, and Davis (2016), and the Google index that is specific to this study. We also compare models estimated over samples of different lengths, since the GI is only available since the first week of January 2004, and an exercise comparing the forecasting performances of models estimated on this sample would be of little practical relevance if models estimated on longer samples were better at predicting the unemployment rate. We also compare the forecasts of Google-based models with those obtained using non-linear models and models based on labor force flows data, as per Barnichon and Nekarda (2012).

We find that Google-based models, estimated using data from 2004 onwards, outperform most of the competitors for predicting the US unemployment rate, irrespective of the length of the time series considered. Their performance improves with the length of the forecast horizon, with the Diebold and Mariano (DM) (1995) test of equal forecast accuracy always rejecting the null at horizons from 1 to 12 months ahead. The analysis of the cumulative sum of squared forecast error differences (CSSED), as suggested by Welch and Goyal (2008), shows that Google-based models perform particularly well during the Great Recession, with their relative performance stabilizing thereafter relative to both the benchmark and the other competing models. Of the various specifications tested, only models using employers' expectations estimated on much longer time series improve on Google-based forecasts at one and two months ahead, though they are outperformed at three to twelve months ahead. We investigate the reasons behind the success of Google-based models for forecasting unemployment further by calculating the transition probabilities at one and twelve months by labor market status and internet job search activity using the Computer and Internet Use supplement of the Current Population Survey. Such estimates suggest that the predictive power of the Google-based indicator is due to the fact that individuals start looking for employment long before losing their job or being re-classified as unemployed instead of inactive. This gives Google-based indicators an advantage over other indicators that provide information on job terminations (initial claims) or employment levels (employers' expectations), but do not cover other relevant elements, such as employees anticipating job loss, unemployment duration, and transitions from inactivity to unemployment.

These results also hold after a number of *robustness checks*, namely: (i) conducting the exercise with the available real time data in the short sample, (ii) employing two alternative, less popular and less relevant, job-search-related keywords, and (iii) conducting a placebo test with a false keyword which is unrelated to job searches but highly correlated with our target variable according to Google Correlate.

We also repeat the forecast horse race for each of the 50 US states plus the District of Columbia (DC) individually, rather than at the federal level, and find that the correlation between the unemployment rate and the Google index is stronger in states in which the percentage of the unemployed who use the internet for job searches is higher. The results of the forecast comparison are less clear cut at the state level than at the federal level, but still point to the substantive forecasting power of Google-based models, which outperform all of their competitors in 35% of the states at one step ahead, and more as the forecast horizon increases (up to a 53% success rate at 12 steps ahead).

Finally, we construct a group of quarterly forecasts of the unemployment rate by combining the monthly forecasts of the best models from our horse race, and compare them with the quarterly predictions released by the Survey of Professional Forecasters (SPF) conducted by the Federal Reserve Bank of Philadelphia. For a given information set, models using the GI outperform the professionals' forecasts, with RMSFEs that are more than 40% lower for nowcasts of the current quarter unemployment rate, and substantially lower for one to three quarters ahead as well. The best Google model also outperforms state-of-the-art models based on labor force flows data, such as that of Barnichon and Nekarda (2012), for three out of four forecast horizons. We also show that Google-based models are quite stable in terms of predictive accuracy.

The innovative data source employed in this article has already been used in epidemiology and in different fields of economics (Edelman, 2012). To the best of our knowledge, the first article using Google data (Ginsberg et al., 2009) involved estimating the weekly '*influenza*' activity in the US using an index of health-seeking behavior that was equal to the incidence of *influenza*-related internet queries. The use of such data in economics started with a paper by Choi and Varian (2012) that showed their relevance for predicting consumer behavior and initial unemployment claims for the US.[2]

To the best of our knowledge, this is the first paper to use Google data to forecast the monthly unemployment rate in the US.[3] However, there has already been some work done for other countries, in particular for Germany

2 Preis, Moat, and Stanley (2013) find evidence that changes in finance-related Google query volumes anticipate stock market moves; Da, Engelberg, and Pengjie (2011) show the relevance of Google data as a direct and timely measure of investors' attention for a sample of Russell 3000 stocks; Baker and Fradkin (forthcoming) develop a job-search activity index in order to analyze the reaction of the job-search intensity to changes in the unemployment benefit duration in the US; Billari, D'Amuri, and Marcucci (forthcoming) use web-search data related to fertility as a leading indicator of the US birth rate; Vosen and Schmidt (2011) show that Google-based indicators improve upon standard ones for forecasting private consumption. Einav and Levin (2013) argue that "big data" (as internet-related data are also called) will have a great impact on economic policy and economic research; see Askitas and Zimmermann (2015) for further discussion and applications.

3 That is, the first other than the previous version of this paper (D'Amuri & Marcucci, 2009). Previously, Ettredge, Gerdes, and Karuga (2005) showed the existence of a positive association between search

(Askitas & Zimmermann, 2009), Italy (D'Amuri, 2009), Israel (Suhoy, 2009), and more recently France (Fondeur & Karamé, 2013). Central banks are also starting to publish reports on the suitability of Google data as a complement to more standard economic indicators (see for example Artola & Galan, 2012; for Spain; McLaren & Shanbhorge, 2011; for the United Kingdom; and Troy, Perera, & Sunner, 2012; for Australia).

Antenucci, Cafarella, Levenstein, Re, and Shapiro (2014) have another interesting application of web-based data to the forecasting of labor market dynamics. They compute an index of job loss, searching and posting using information made available via Twitter accounts, and show that such an indicator has the potential to improve predictions of initial claims dynamics for the US. Twitter-based data have the clear advantage that they make it possible to track individuals over time and to correlate their activity with self-declared personal characteristics. At the same time, they suffer more from sample selection for two reasons: (i) the use of Twitter is less widespread than Google searches, and (ii) individuals are not always willing to share personal information about their job market status on social media, whereas Google searches are anonymous.

Based on our results for the unemployment rate, we believe that there is the potential for further applications of Google query data to other fields of economics.

The paper is organized as follows. Section 2 describes the data used to predict the US unemployment rate, with a particular emphasis on the GI. Section 3 discusses the models that we employ for predicting the US unemployment rate, while Section 4 compares their out-of-sample performances. Section 5 conducts a number of robustness checks. Section 6 compares our predictions with those of the Survey of Professional Forecasters and with those of models based on labor force flows, while Section 7 concludes.

2. Data and descriptive statistics

The data used in this paper come from various different sources. The seasonally adjusted monthly unemployment rate[4] is the one released by the Bureau of Labor Statistics (BLS), and comes from the current employment statistics and the local area unemployment statistics for the national and state levels, respectively.[5]

We complement these data with a well-known leading indicator for the unemployment rate (see for example Montgomery et al., 1998): the weekly seasonally-adjusted IC released by the US Department of Labor.[6]

We also consider alternative leading indicators that are used routinely by economic and financial observers in forming their expectations on labor market prospects, but whose power in predicting unemployment has never been assessed before in the economic literature, to the best of our knowledge. Specifically, our exogenous variables are employment expectations for the manufacturing and non-manufacturing sectors from the Institute for Supply Management's (ISM) Report on Business (EEM_t and $EENM_t$ respectively), the current and six-month-ahead consumer expectations from the US Consumer Confidence survey of the Conference Board (CE_t and $CE6M_t$ respectively), and the index of economic policy uncertainty proposed by Baker et al. (2016).[7]

2.1. Google-based data

The exogenous variable that is specific to this study is the weekly GI that summarizes the job searches performed through the Google search engine website, collected through Google Trends. The GI represents the number of web searches that have been made for a particular keyword in a given geographical area (r) within a given time. The search share for a particular keyword on day d is given by the number of web searches containing that keyword ($V_{d,r}$), *normalized* by division by the total number of web searches performed through Google for the same day and area ($T_{d,r}$), i.e., $S_{d,r} = \frac{V_{d,r}}{T_{d,r}}$. The search share for week τ is given by the simple average $S_{\tau,r} = \frac{1}{7}\sum_{d=Sunday}^{Saturday} S_{d,r}$. For privacy and anonymity reasons, no

engine keyword usage data extracted from WordTracker's Top 500 Keyword Report and the monthly number of unemployed for the interval between September 2001 and March 2003. More recently, Tuhkuri (2016) showed that data on Google searches can sometimes help to predict the US unemployment rate, especially at short horizons. Nevertheless, Tuhkuri's (2016) results are not comparable to ours for the following reasons: (i) the web-search-based leading indicator used only tracks the interest in unemployment benefits, but some of the unemployed are not eligible for unemployment insurance (for example, individuals who are looking for their first job), and some of those who are eligible do not claim it (e.g., the take-up rate in the US around the Great Recession was about 45% according to East & Kuka, 2015); (ii) the author adopts data that have not been adjusted for seasonality, using a seasonal factor for the target variable only; and (iii) he does not align the weeks used to compute the monthly Google index with the reference weeks used by the BLS to calculate the official unemployment rate (see Section 2.1).

4 The unemployment rate for month t refers to individuals who do not have a job, but are available for work, in the week including the 12th day of month t (i.e., the reference week), and who have looked for a job in the four weeks prior to the reference week. The monthly unemployment rate is released on the first Friday of month $t+1$.

5 Many papers in the literature impose the presence of a unit root or induce stationarity with a particular transformation (see for example Rothman, 1998). However, Montgomery, Zarnowitz, Tsay, and Tiao (1998) model the level of the monthly unemployment rate, arguing that it is hard to justify unit-root non-stationarity for the US unemployment rate because it is a rate that varies within a limited range. Similarly, Koop and Potter (1999) argue that the unemployment rate cannot exhibit a global unit root behavior, since it is bounded by 0 and 1. Previous versions of this paper tested for the presence of a unit-root formally, using tests that are robust to non-linearities and structural breaks, and found opposite results for the short and long samples. Thus, we have adopted the more agnostic approach of Koop and Potter (1999) and Montgomery et al. (1998), deciding not to restrict our models to the stationary regime explicitly, and presenting all of our forecasting results using the levels of the monthly US unemployment rate.

6 Since seasonally adjusted data are issued only at the national level, we have performed our own seasonal adjustment for the state-level data using the X-13 ARIMA-SEATS filtering of the US Census Bureau. Regarding the timing, IC data for the jth week of month t ($wj(t)$) are released by the Department of Labor on Thursday of the $(j+1)$th week.

7 We also considered the alternative uncertainty indices suggested by Jurado, Ludvigson, and Ng (2015) and Orlik and Veldkamp (2014), but they had rather low correlations with both the unemployment rate and the GI (Table A.3 of the Appendix), and thus we did not include them among the exogenous variables used in the horse race.

absolute values of the GI components are available publicly. Google also scales the index $GI_{t,r}$ to 100 in the week in which it reaches the maximum level. Thus, the Google index for week τ is given by $GI_{\tau,r} = \frac{100}{\max_{\tau}(S_{\tau,r})} S_{\tau,r}$, and represents the likelihood of a random user from that area doing a Google search for that particular keyword during that week. The data are gathered using IP addresses, and are made available to the public if the number of searches exceeds a certain – undeclared – threshold. Repeated queries from a single IP address within a short period of time are eliminated. The data are updated weekly and are available almost in real time starting with the week ending January 10, 2004. Google trends data are freely available at https://www.google.com/trends/?hl=en. The user-friendly interface permits users to download time series for up to five keywords or combinations of keywords for a given country, or the same keyword for up to five countries. Once the data have been retrieved, they can be downloaded in csv format provided that the user has a Google account. Last but not least, the index is calculated based on a sample of IPs that changes with time (daily) and with the IP address. As a consequence, the indices can vary according to the day and the IPs of download. Throughout the paper, we compute our indices as the simple average of 24 downloads carried out over 12 different days from two different IPs. Nevertheless, taking the raw data coming from the single downloads would not alter the results much, since the elementary time series are nearly identical, with cross-correlations that are never below 0.99.[8]

Our preferred indicator summarizes the incidence of queries that include the keyword "*jobs*" out of all queries performed through Google in the relevant week (this index is labeled $G1$ henceforth).[9] We choose to use the keyword "*jobs*" as the main indicator of job-search activities for two main reasons. First, we found the keyword "*jobs*" to be the most popular among different job-search-related keywords. No absolute search volumes are available, but it is possible to identify the most popular keywords by looking at relative incidences. Figure A.1 of the online Appendix plots the monthly average values of the GI for the keywords "*facebook*", "*youtube*", and "*jobs*". Of these keywords, "*facebook*" has the highest incidence, while the GI for "jobs" is constantly around 10. This means that, when searches for "*facebook*" were at their peak, there was still one keyword search for "*jobs*" for every 10 searches for "*facebook*". The results are similar when conducting the comparison with the keyword "*youtube*", another popular search, which reaches a maximum level of above 40 in our sample.

In addition to its popularity, the second reason why we chose the keyword "*jobs*" is that we believe that it is used most widely across the broadest range of job seekers, and therefore is less sensitive to the presence of demand or supply shocks that are specific to subgroups of workers, which could bias the value of the GI and its ability to predict the overall unemployment rate. Finally, it has to be noted that the numerator of the index contains all of the keyword searches that include the word "*jobs*", such as "*public jobs*" or "*California jobs*", for example. As a consequence, the index is based on a broader set of queries that include the word "*jobs*", some of which might actually be unrelated to job searches. Such a measurement error is unlikely to be correlated with the monthly unemployment rate over time and, if anything, should reduce the predictive power of our leading indicator.[10]

The variable also has other limitations. For example, individuals who are looking for a job through the internet (jobs available through the internet) may well be not selected randomly among job seekers (jobs). Moreover, the indicator captures overall job-search activities; that is, the sum of searches performed by unemployed and employed people. This limitation is made more severe by the fact that, while unemployed job searches are believed to follow the anti-cyclical variation of job separation rates, on-the-job searches are normally assumed to be cyclical. We acknowledge that this could introduce some bias into our GI; nevertheless, if anything, such a bias should reduce the precision of our forecasts based on Google data.

We should also consider the representativeness of internet data in general, and of Google data in particular. According to the July supplement to the 2011 Current Population Survey, 30.1% of those unemployed use the internet to look for a job. Moreover, according to comScore, Google persistently had a dominant share of the search engine market between 2004 and 2014, going from 56% in 2004 to 67% in 2014. Thus, these data have the potential to track social phenomena if there is a connection between what people search for on the web and their later behavior.

Our empirical analysis aligns the GI and IC data with the relevant weeks for the unemployment survey.[11] Figs. 1 and 2 depict the unemployment rate, along with all of the exogenous variables evaluated in this article.[12] IC and the GI are highly correlated with the level of the unemployment rate: 0.64 for IC and 0.8 for the GI; consumers' and employers' expectations show lower

8 The online Appendix (see Appendix A) reports graphs (Figure A.7 and B.1 to B.5), descriptive statistics (Table A.4) and correlations (Table A.5) for the 24 raw time series. We have also computed the forecasting results using all of the individual GIs downloaded at each IP address and each day, which are summarized for the G1 monthly averages in Figure A.8 of the online Appendix. The other results are very similar and are available from the authors upon request.

9 We have adjusted both the weekly and monthly indicators for seasonality using the X13 ARIMA-SEATS filter. Google data have a peculiar seasonality: there are usually troughs in November and December when the denominator of the index gets inflated by Christmas-related searches.

10 However, we do subtract from the numerator the keyword searches for "*Steve Jobs*", a popular search that includes the word "*jobs*", in order improve the precision of our GI.

11 When constructing the GI or the IC for month t, we take into account the week that includes the 12th of the month and the three preceding weeks, which is the exact same interval that is used to calculate the unemployment rate for month t, reported in the official statistics (what we may call the 'survey time'). When there are more than four weeks between the reference week of month t and the following one in month $t+1$, we do not use either the GI or the IC for the week that is not used by the official statistics (the first week after the reference week of month t) to calculate the unemployment rate (see Figure A.2 of the online Appendix for a visual description of the alignment procedure and for the timing of our variables).

12 Table A.1 of the online Appendix reports the descriptive statistics for the monthly US unemployment rate and various leading indicators for the interval 2004.1–2014.2.

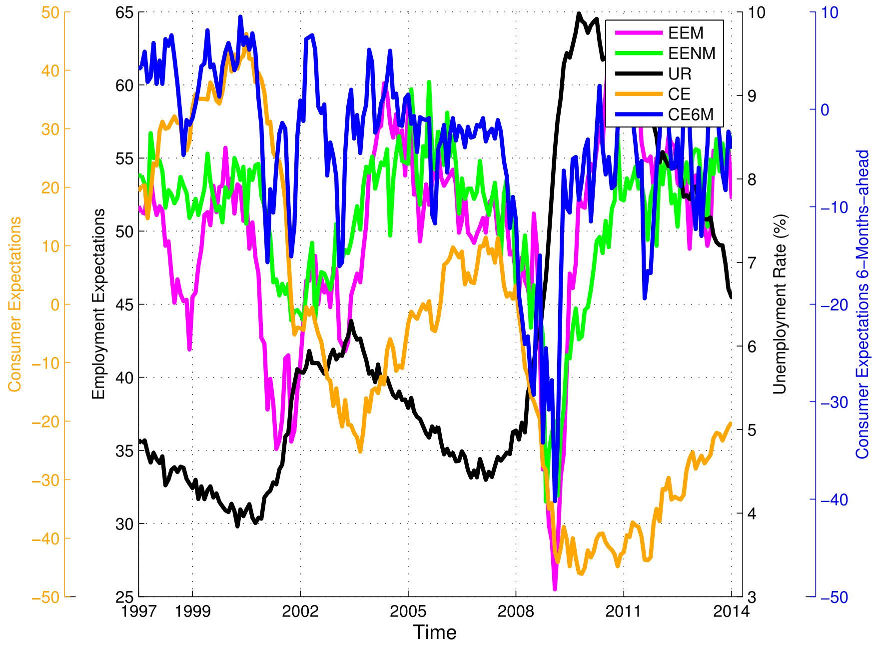

Fig. 1. US unemployment rate and leading indicators: employers' and consumers' expectations. *Notes*: The employment expectations for the manufacturing (EEM) and non-manufacturing (EENM) sectors are from the ISM Report on Business, while the current (CE) and six months in advance (CE6M) consumer expectations are from the US Consumer Confidence Survey of the Conference Board.

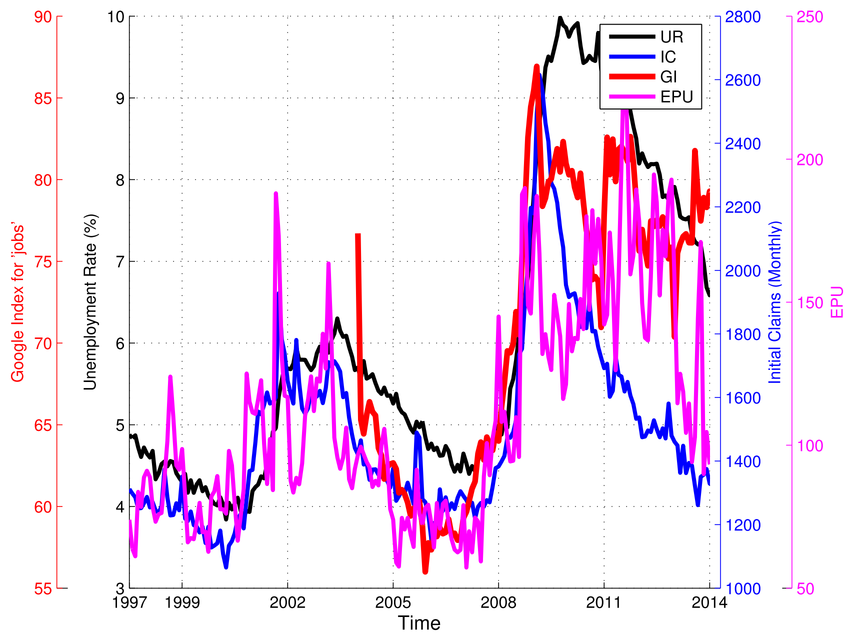

Fig. 2. US unemployment rate and leading indicators: initial claims, EPU, and the Google index. *Notes*: The initial claims (IC) are monthly averages of the weekly IC. The Google index (GI) is the monthly average of Google 'jobs' searches. EPU is the economic policy uncertainty indicator.

correlations (Table A.2 of the online Appendix). Thus, the correlations of the GI for "*jobs*" with the unemployment rate are higher than those of the IC, which is the most widely accepted leading indicator in the literature.

3. Forecasting models and methods

Our paper focuses on the multi-step pseudo out-of-sample forecasting performances of a variety of linear models for nowcasting and forecasting the US unemployment rate. Following the literature on forecasting economic variables (Stock & Watson, 2003), we adopt a standard autoregressive model with explanatory variables as follows:

$$y^h_{t+h} = \beta_0 + \beta_1(L)y_t + \beta_2(L)x_t + \varepsilon_{t+h},\quad t = 1, 2, \ldots, T, \tag{1}$$

where y^h_{t+h} is the h-period-ahead unemployment rate at time t; x_t is a possible explanatory variable or leading indicator in period t; y_t is the period t unemployment rate; and ε_{t+h} is an error term. $\beta_1(L)$ and $\beta_2(L)$ are lag polynomials, such that $\beta_1(L)y_t = \sum_{j=1}^{p} \beta_{1j}y_{t-j}$ and $\beta_2(L)x_t = \sum_{j=1}^{q} \beta_{2j}x_{t-j}$. The lag orders p and q are selected recursively and sequentially at each forecast origin using the BIC.[13] We consider from one- to twelve-step-ahead *direct* forecasts of the US unemployment rate by setting the parameter $h = 1, 2, \ldots, 12$ months.

We compare bivariate models which differ only in the additional explanatory variable x_t that is adopted for forecasting the target variable. As possible explanatory variables, we adopt a series of leading indicators for the unemployment rate. We begin with the most commonly used indicator, initial claims (IC), and use either its monthly average or its weekly value. In addition, we also evaluate the economic policy uncertainty (EPU) index, the employment expectations for the manufacturing (EEM) and non-manufacturing (EENM) sectors, plus current and six-month-ahead consumer expectations (CE and CE6M, respectively). Finally, we consider the GI for the keyword "*jobs*", as was explained in the previous section. In this case, we add as an explanatory variable either its monthly average or its weekly values.

We compare the h-step-ahead pseudo out-of-sample forecasting performances of these models with that of a univariate autoregression ($AR(p)$), where the lag p is selected recursively by the BIC. We refer to the latter model as our benchmark model:

$$y^h_{t+h} = \beta_0 + \beta_1(L)y_t + \eta_{t+h},\quad t = 1, 2, \ldots, T. \tag{2}$$

Eqs. (1) and (2) are both estimated by OLS, in rolling samples of two different lengths. In the short sample, they are estimated over a rolling window of 37 observations ($R = 37$) starting from February 2004.[14] We also use a longer time interval starting in 1997:7, the first month in which all of the non-Google-based exogenous variables used in this article are available. In the latter case, Eqs. (1) and (2) are estimated over a rolling window of 116 observations ($R = 116$).[15] Accordingly, the first 1-month-ahead out-of-sample forecast is made for 2007:3, while the first 12-month-ahead out-of-sample forecast is made for 2008:2.

Finally, the exercise is carried out with both revised and (for the short sample only) real time data.[16]

4. Out-of-sample forecasting comparison

4.1. Methodology

We evaluate the out-of-sample performances of our competing models relative to the benchmark by comparing the root mean squared forecast error (RMSFE) of each model with that of the benchmark, and we test for equal forecast accuracy using a Diebold and Mariano (1995, DM) test. Moreover, we also compare forecasts based on the cumulative sum of squared forecast error differences (CSSED), introduced by Welch and Goyal (2008), to check for possible forecast instabilities.

4.2. RMSFE comparison

Panel A of Table 1 reports the results of the forecast comparison over the short sample 2004.2–2014.2 (first in-sample 2004.2–2007.2), for forecast horizons of 1 to 12 steps ahead. The first row reports the root mean squared forecast error (RMSFE) of the benchmark $AR(p)$ model estimated over the short sample. For all other rows, the values reported are the ratios of the RMSFE of the model using the leading indicator in each row to that of the benchmark model. A value below one means that the model in the row has a lower RMSFE than the benchmark (and thus outperforms it), and *vice versa*. Most of the competing models tend to outperform the benchmark at short forecast horizons. The models using initial claims, the standard leading indicator for the unemployment rate, outperform the benchmark at short term horizons (up to four months ahead), but do not perform particularly well at longer horizons. Alternative models that include the economic policy uncertainty (EPU) index of Baker et al. (2016), employment or consumer expectations tend to fare relatively better. This is true in particular for models that include *EENM* (employment expectations in the non-manufacturing sectors) or *CE* (consumer expectations on current labor market dynamics); these are able to reduce the RMSFE compared to the benchmark at all forecast horizons, with the results for *EENM* leading to a rejection of the null of equal forecast accuracy (DM test) for forecasts from one to seven steps ahead. Moving to Google-based models, those using the monthly means of the GI achieve the best performances in terms of RMSFEs, with an advantage over the benchmark that actually increases with the forecast horizon (18% improvement at one step ahead, 32% at 12 steps ahead). The DM test rejects the null of equal forecast accuracy compared to the benchmark at any

13 First, the lag p is selected based on the in-sample estimation of the benchmark model $AR(p)$ in (2). After choosing the "best" benchmark specification, the lag q is chosen based on the in-sample estimation of Eq. (1). The maximum lag length considered in both cases is four. Table A.6 in the online Appendix shows the average in-sample BICs for all of the competing models at each forecast horizon.

14 We start from February 2004 because data for $G1_{W1,t}$ and $G1_{W2,t}$ (the weekly indices) are not available for January 2004.

15 We restrict the analysis to the period 1997:7 onward due to the availability of employment expectations in the non-manufacturing sector.

16 For the initial claims data, we combine the real time data available on the Department of Labor website with those available from the Fred database (the latter reports real time data starting from May 28, 2009). For the unemployment rate, we use Bureau of Labor Statistics data.

Table 1
Out-of-sample results for the US unemployment rate in levels (u_t): rolling scheme, direct forecasts, GI for "jobs".

x_{it}	$h = 1$	$h = 2$	$h = 3$	$h = 4$	$h = 5$	$h = 6$	$h = 7$	$h = 8$	$h = 9$	$h = 10$	$h = 11$	$h = 12$
Panel A: Short sample												
$AR(p) - SS(RMSFE)$	0.206	0.319	0.454	0.561	0.645	0.760	0.877	1.000	1.108	1.209	1.322	1.358
IC_t	0.847**	0.875**	0.836***	0.888**	0.941	0.975	0.990	1.012	1.002	0.992	0.982	1.024
$IC_{W1,t}$	0.891	0.932	0.918	0.946	0.962	0.969	0.984	1.000	1.013	1.014	1.023	1.023
$IC_{W2,t}$	0.884**	1.026	1.026	1.061	1.019	1.017	1.026	1.000	1.009	1.015	1.010	1.008
$IC_{W3,t}$	0.875**	0.912	0.899	0.972	0.983	1.010	1.026	1.014	1.059	1.063	1.077	1.097
$IC_{W4,t}$	0.835*	0.795*	0.781*	0.898	0.974	1.020	1.053	1.073	1.048	1.060	1.049	1.054
EEM_t	0.859*	0.887	0.877	0.932	1.007	1.041	1.075	1.061	1.068	1.101	1.088	1.142
$EENM_t$	0.824**	0.777**	0.746**	0.751***	0.797**	0.818**	0.834*	0.848	0.845	0.842	0.831	0.867
CE_t	0.889	0.858	0.771	0.765	0.755	0.756	0.742	0.719	0.712	0.718	0.710	0.700
CE_t^{6M}	0.894	0.873	0.805	0.811	0.826	0.801	0.760	0.734	0.722	0.719	0.706	0.754
EPU_t	0.969	0.918	0.884	0.846	0.870	0.852	0.927	0.944	0.972	0.968	0.944	0.950
$G1_t$	*0.821**	*0.765**	**0.656***	**0.652***	**0.644***	**0.662***	**0.659***	**0.665***	**0.680***	**0.678***	**0.673***	**0.685***
$G1_{W1,t}$	0.952	0.883	0.967	1.066	1.195	1.276	1.311	1.287	1.252	1.225	1.204	1.231
$G1_{W2,t}$	0.901	0.975	0.967	1.006	1.086	1.051	1.021	1.008	0.971	0.989	0.980	0.907
$G1_{W3,t}$	0.866*	0.802*	0.682*	0.687*	0.704*	0.710*	0.700*	0.682*	0.683*	0.709*	0.731*	0.766*
$G1_{W4,t}$	0.843**	0.777*	0.672*	0.694*	0.691*	0.716**	0.798**	0.798**	0.813*	0.798*	0.793*	0.808*
Panel B: Long sample												
$AR(p) - LS$	0.891**	0.905*	0.839*	0.848*	0.871*	0.877**	0.880**	0.896**	0.909*	0.936*	0.949*	1.009+
$IC_t - LS$	0.837*	0.830	0.817	0.851	0.898	0.925	0.927	0.947	0.960	1.001	1.000	1.061
$IC_{W1,t} - LS$	0.860	0.855	0.844	0.871	0.906	0.922	0.931	0.940	0.953	0.984	0.995	1.055
$IC_{W2,t} - LS$	0.861	0.871	0.838	0.864	0.901	0.912	0.921	0.935	0.951	0.980	0.990	1.049
$IC_{W3,t} - LS$	0.843*	0.844	0.817	0.853	0.896	0.914	0.924	0.942	0.956	0.987	0.998	1.057
$IC_{W4,t} - LS$	0.809**	0.774*	0.761	0.823	0.872	0.895	0.904	0.916	0.937	0.973	0.992	1.051
$EEM_t - LS$	0.828**	0.833	0.812	0.846	0.879	0.906	0.916	0.920	0.934	0.957	0.965	1.031
$EENM_t - LS$	**0.802****	**0.746***	*0.720*	*0.742*	*0.770*	0.784	0.806	0.824	0.846	0.866	0.877	0.943
$CE_t - LS$	0.886	0.887	0.827	0.844	0.864	0.862	0.859	0.885	0.897	0.928	0.942	1.003
$CE_t^{6M} - LS$	0.821**	0.815	0.788	0.772	0.776	*0.750*	*0.743*	*0.736*	*0.748*	*0.768*	*0.782*	*0.834*
$EPU_t - LS$	0.901	0.929	0.869	0.875	0.894	0.890	0.888	0.904	0.920	0.945	0.955	1.021
$SETAR(2) - LS$	0.978	1.370+	1.497	1.650	1.827+	1.879+	1.881+	1.842+	1.858+	1.865+	1.818++	1.886++
$LSTAR(2) - LS$	0.988	1.438++	1.515+	1.767+	1.901+	1.942+	1.935+	1.878+	1.864++	1.869++	1.814++	1.850++
$AAR(2) - LS$	0.965	1.317+	1.412	1.620	1.770	1.818	1.839+	1.815+	1.824+	1.835+	1.799+	1.866+

Notes: The short sample is 2004.2–2014.2, while the long sample is 1997.7–2014.2; short in-sample: 2004.2–2007.2; long in-sample: 1997.7–2007.2; out of sample: 2007.3–2014.2. For the benchmark model $AR(p)$, the RMSFE is indicated, while for all of the other models the table reports the ratio of the competing model's RMSFE in each row to that of the $AR(p)$ benchmark model estimated in the short sample. Thus, a number below 1 indicates that the comping model beats the $AR(p)$ benchmark, while a number above 1 means that the benchmark performs better. For each panel, numbers in boldface correspond to the overall lowest RMSFE at each forecast horizon (h), while numbers in italics are the lowest RMSFE for each panel. In all panels, *** (+++), ** (++) and * (+) indicate rejection at the 1, 5 and 10% significance levels, respectively, of the Diebold and Mariano (DM, 1995) test for the null of equal forecast accuracy between the $AR(p)$ benchmark estimated over the short sample and the AR model with the alternative leading indicator indicated in the first column when the benchmark is beaten by (beats) the competitor model.

forecast horizon. Slightly worse results are obtained when using the weekly values of the indicator for the last two weeks of the BLS survey reference period, while models that use the GI for the first two weeks do not outperform the benchmark.

Panel B of Table 1 also checks whether models that include initial claims, the EPU index or employment expectations estimated on a longer time interval (starting in 1997.7, due to the availability of employers' expectations in the non-manufacturing sector) for which Google data were not available can provide more accurate forecasts. The table shows that forecasting performances do not change significantly when using this longer interval. Models that use employment expectations for the non-manufacturing sector, which already performed well in the short sample, improve their forecast accuracy further and beat models that use GI by a little margin, but only for one- and two-step-ahead forecasts.

Finally, we also estimate three non-linear time series models[17]: the self-exciting threshold autoregression (SETAR) model with two lags, the logistic smooth transition autoregressive (LSTAR) model with two lags and the additive autoregressive model (AAR) with two lags. In no case were these models able to outperform the benchmark.

Moreover, these patterns are confirmed when we repeat the exercise on the short sample using real time data with no revisions (see Table 2). In this case, the initial claims tend not to perform particularly well, while employer and consumer expectations and the EPU index improve relative to the benchmark. Nevertheless, the best leading indicator remains the one that includes the monthly Google index for "*jobs*", outperforming all

[17] See Golan and Perloff (2004), Hastie and Tibshirani (1990), Montgomery et al. (1998) and Rothman (1998) for previous applications.

Table 2
Out-of-sample results for the US unemployment rate in levels (u_t) in real time: rolling scheme, direct forecasts, GI for "jobs".

x_{it}	$h=1$	$h=2$	$h=3$	$h=4$	$h=5$	$h=6$	$h=7$	$h=8$	$h=9$	$h=10$	$h=11$	$h=12$
$AR(p)$ ($RMSFE$)	0.22	0.33	0.47	0.55	0.66	0.78	0.87	0.97	1.00	1.06	1.17	1.29
IC_t	0.95	0.91	0.89**	0.93	0.99	0.92*	0.98	1.05	1.05	1.06	1.05	1.03
$IC_{W1,t}$	0.93	0.94	0.95	0.96	0.97	0.94	0.97	1.01	1.02	1.04	1.04	1.04
$IC_{W2,t}$	0.94	0.98	1.00	1.08	1.03	1.01	1.04	1.05	1.09	1.11	1.13	1.11
$IC_{W3,t}$	1.00	0.97	0.97	1.01	0.98	0.94	1.00	1.01	1.06	1.09	1.05	1.03
$IC_{W4,t}$	0.98	0.86	0.85	0.96	0.95	0.91	0.98	0.97	0.98	1.02	1.00	1.01
EEM_t	0.98	0.91	0.90	0.91	0.92	0.95	1.01	1.06	1.12	1.18	1.19	1.20
$EENM_t$	0.90	0.80*	0.74**	0.74***	0.74***	0.75**	0.81**	0.85*	0.93	0.89	0.97	0.95
CE_t	0.91	0.86	0.78	0.80	0.72*	0.70*	0.74*	0.72	0.77	0.78	0.79	0.76
CE_t^{6M}	0.93	0.86	0.80	0.84	0.78*	0.75*	0.74**	0.74*	0.84	0.85	0.84	0.86
EPU_t	0.99	0.92	0.90	0.89	0.87	0.87	0.88	0.90	0.95	0.93	0.94	0.93
$G1_t$	0.89*	**0.76***	**0.65****	**0.66****	**0.61****	**0.61****	**0.63****	**0.67***	0.74*	**0.75***	**0.77****	**0.75****
$G1_{W1,t}$	0.98	0.88	0.85	0.86	0.99	1.10	1.22	1.25	1.35	1.39	1.40	1.35
$G1_{W2,t}$	**0.87****	0.86*	0.82	0.90	1.02	0.95	0.95	1.03	1.09	1.14	1.07	1.02
$G1_{W3,t}$	**0.87***	0.80*	0.68*	0.68**	0.66**	0.67**	0.68**	0.68*	**0.73***	0.77*	0.79**	0.80*
$G1_{W4,t}$	**0.87****	0.78*	0.69*	0.71*	0.64**	0.66**	0.75**	0.80*	0.86	0.87	0.86*	0.85*

Notes: The short sample is 2004.2–2014.2, with the in-sample period being 2004.2–2007.2 ($R = 37$ observations) and the out-of-sample period being 2007.3–2014.2. For the benchmark model $AR(p)$, the root MSFE (RMSFE) is reported, while for all of the other models the table displays the ratios of the RMSFEs of the competing models in each row to that of the benchmark. Thus, a number below 1 indicates that the competitor model beats the $AR(p)$ benchmark, while a number above 1 indicates that the benchmark performs better. Numbers in boldface correspond to the lowest RMSFE for each forecast horizon (h).
In all panels ***, ** and * indicate rejection at the 1, 5 and 10% significance levels, respectively, of the Diebold and Mariano (DM, 1995) test for the null of equal forecast accuracy between the AR(p) benchmark estimated over the short sample and the AR model with the alternative leading indicator indicated in the first column.

Table 3
Year 2011: Transition matrix by internet job search activity.

Status in month T	Internet job search	Status in $T+1$			Status in $T+12$		
		E	U	I	E	U	I
Employed (E)	No	95.8	1.3	2.8	91.9	3.1	5.0
Employed (E)	Yes	96.5	1.9	1.6	92.3	3.7	4.0
Unemployed (U)	No	18.4	58.1	23.5	42.7	29.9	27.4
Unemployed (U)	Yes	16.2	70.3	13.6	50.1	33.1	16.8
Inactive (I)	No	5.8	5.1	89.1	18.2	7.0	74.8
Inactive (I)	Yes	8.6	15.4	76.0	25.4	15.3	59.3

Notes: The transition probabilities are estimated from the IPUMS-Current Population Survey data panel (King et al., 2010). The results for years in which an Internet and Computer Use Supplement is present in the CPS survey are reported in Table A.7 of the online appendix.

competing models for 10 of the 12 forecast horizons (the best models in the other two cases include weekly GI values).

These results highlight the predictive power of leading indicators based on Google data, with alternative indicators such as employment expectations beating them only when forecasting at short horizons (one and two months ahead).

We obtain an understanding of the determinants of such predictive power by estimating transition probabilities by initial labor force status separately for individuals who did and did not search for jobs online. In particular, we start from a month in which the Supplement on Internet and Computer Use of the Current Population Survey asks respondents whether they have carried out online job searches, then follow these individuals over time using the panel component of the survey (King et al., 2010) and collect their labor force statuses one and twelve months later. The results reported in Table 3 show that individuals who were employed in July 2011 but were conducting internet job searches were more likely to be unemployed one and twelve months later than employed individuals who did not conduct such searches. This higher incidence of unemployment among individuals who are looking for a job online is found also for those who were unemployed or inactive, and when estimating the same transitions for all years in which the Computer and Internet use Supplement is present in the Current Population Survey (Table A.7 of the online Appendix). These results suggest that the good forecasting performance of the Google-based indicator can be attributed to the fact that individuals start to look for employment long before losing their job or being reclassified as unemployed instead of inactive. This gives Google-based indicators an advantage over the standard initial claims, which only capture contemporaneous job terminations and do not provide information on other relevant margins such as unemployment duration or transitions from inactivity to unemployment in either the current month or the months ahead.

4.3. Possible forecast instabilities and the CSSED evolution

One possible concern regarding the use of a single model for forecasting over long time periods is that the predictive accuracy might depend on the particular out-of-sample period used for the forecast evaluation. In particular, a model might be chosen for having the highest predictive accuracy when the loss functions are evaluated over the whole out-of-sample period, but one of the competing models might exhibit a lower RMSFE at a particular point (or points) in time over the evaluation period. Moreover, the nature of Google-based data means that it intrinsically evolves over time, due to the changing group of people who use the internet (and Google) for job searches over time, and also to changes in technology (such as the spread of smartphones). Such an evolution could have a non-negligible impact on the forecast comparison.

This section checks whether the superior forecast performance of Google-based models is due to particular out-of-sample sub-intervals by reporting the cumulative sum of squared forecast error difference (CSSED) values for the best models using IC, expectations, EPU and GI across forecast horizons (Fig. 3).[18] At any point in time, a CSSED value above zero means that the alternative model outperforms the benchmark, while positive (negative) changes in the slope of the CSSED suggest that there is an increase (decrease) in the relative performance of the competing model with respect to the benchmark.

The movements of CSSED over time confirm that the model using GI is the best throughout the evaluation sample. It is clear that models that use either the GI or expectations definitely outperform the AR benchmark at all forecast horizons during the Great Recession (the slope is increasing and steep), while the model using IC outperforms the benchmark only at the shortest forecast horizons (until $h = 4$). In the period following the Great Recession, all models tend to stabilize in terms of performances relative to the benchmark.[19]

5. Robustness checks

This section performs the following *robustness checks* for the main results presented thus far: (i) we test the forecasting properties of two alternative, less popular and less relevant, job-search-related keywords, (ii) we perform a placebo test, (iii) we repeat the forecast horse race for each of the 50 US states plus DC individually, rather than at the federal level. All of these tests confirm, either directly or indirectly, the good performance of Google-augmented models for forecasting the monthly US unemployment rate.

5.1. Alternative keywords

As a first robustness check, we analyze the properties of our forecasting models using not only our preferred GI for "*jobs*" (G1), but also other keywords that are related closely to job searches. In particular, we look at the Google indices for "*job center*" (labeled G2) and the sum of these two keywords and "*collect unemployment*", labeled G4.

Figure A.5 of the online Appendix plots the dynamics of the alternative GIs together with the monthly US unemployment rate; a visual inspection reveals these two alternative leading indicators to have patterns similar to those of the time series we are forecasting. The two indices are correlated very highly with the contemporaneous unemployment rate (0.84 and 0.86, respectively).[20]

Table 4 carries out a horse race across these three different indices. The results for G1 are obviously the same as those in the main exercise (Table 1). The forecasts obtained using the alternative and less-representative keyword "*job center*" (G2) are very similar to those obtained by the benchmark AR(p) model at any forecast horizon. The situation improves when the G4 index is used, with the forecasts beating the benchmark at every forecast horizon and the relative forecast performance increasing steadily with the forecast horizon. Nevertheless, the G1-based forecasts remain superior to those made using these alternative indices at all forecast horizons.[21]

5.2. Placebo test

This section provides a placebo test for the main results of this paper. In particular, we test the forecasting power of an alternative Google-based indicator, which is chosen to be the one with the highest correlation with the time series of the monthly US unemployment rate in the estimation sample, but is not necessarily related to job searches. This keyword can be selected thanks to an application developed by Google, called 'Google Correlate',[22] that is able to identify, within a specified time interval, the keyword web searches that (i) show the highest correlations with a given keyword search, and (ii) show the highest correlations with a given time series. In particular, we isolated the time series of the US monthly unemployment rate and used this application to find the set of keyword searches that, of all of the searches conducted through the search engine, were correlated with it most closely in the first in-sample period (2004.2–2007.2). Of these, we selected the GI for the keyword '*dos*', which has no logical connection to job searches:

18 CSSED is defined as $CSSED_{m,\tau} = \sum_{\tau=R}^{T}(\hat{e}^2_{bm,\tau} - \hat{e}^2_{m,\tau})$, where $\hat{e}^2_{bm,\tau}$ denotes the squared forecast error of the AR benchmark model and $\hat{e}^2_{m,\tau}$ denotes the same for the competing model. R and T are the beginning and end of the forecast evaluation period, respectively.

19 We investigate the stability of the forecast accuracies of our Google-augmented models further by computing RMSFE ratios with respect to the benchmark $AR(p)$ over 46 rolling sub-samples. Figure B.10 to B.12 of the online Appendix depict these ratios for all models using IC, expectations or the EPU index, or GI for "*jobs*" at each horizon. Note that models that use IC tend to have higher predictive accuracies throughout the evaluation sample for up to three months ahead, but for longer forecast horizons they outperform the benchmark only after the Great Recession. Models that use expectations or the EPU index all tend to outperform the benchmark except for the model using *EEM*, as do all models that adopt the GI. The only exception is for the models that use the GI for the first and second weeks of the survey, which perform poorly, especially in the first part of our evaluation period. Note that the forecasting performances of models that use GI tend to be similar to those of the AR benchmark at the end of the sample.

20 The descriptive statistics for the two indices, at both the monthly and weekly levels, are reported in Table A.1 and A.3 of the online Appendix.

21 Figure B.13 and B.15 of the online Appendix show that the forecasts from these alternative GIs are quite unstable and definitely perform poorly relative to G1.

22 Available at www.google.com/trends/correlate/. See Mohebbi et al. (2011) for details of this application.

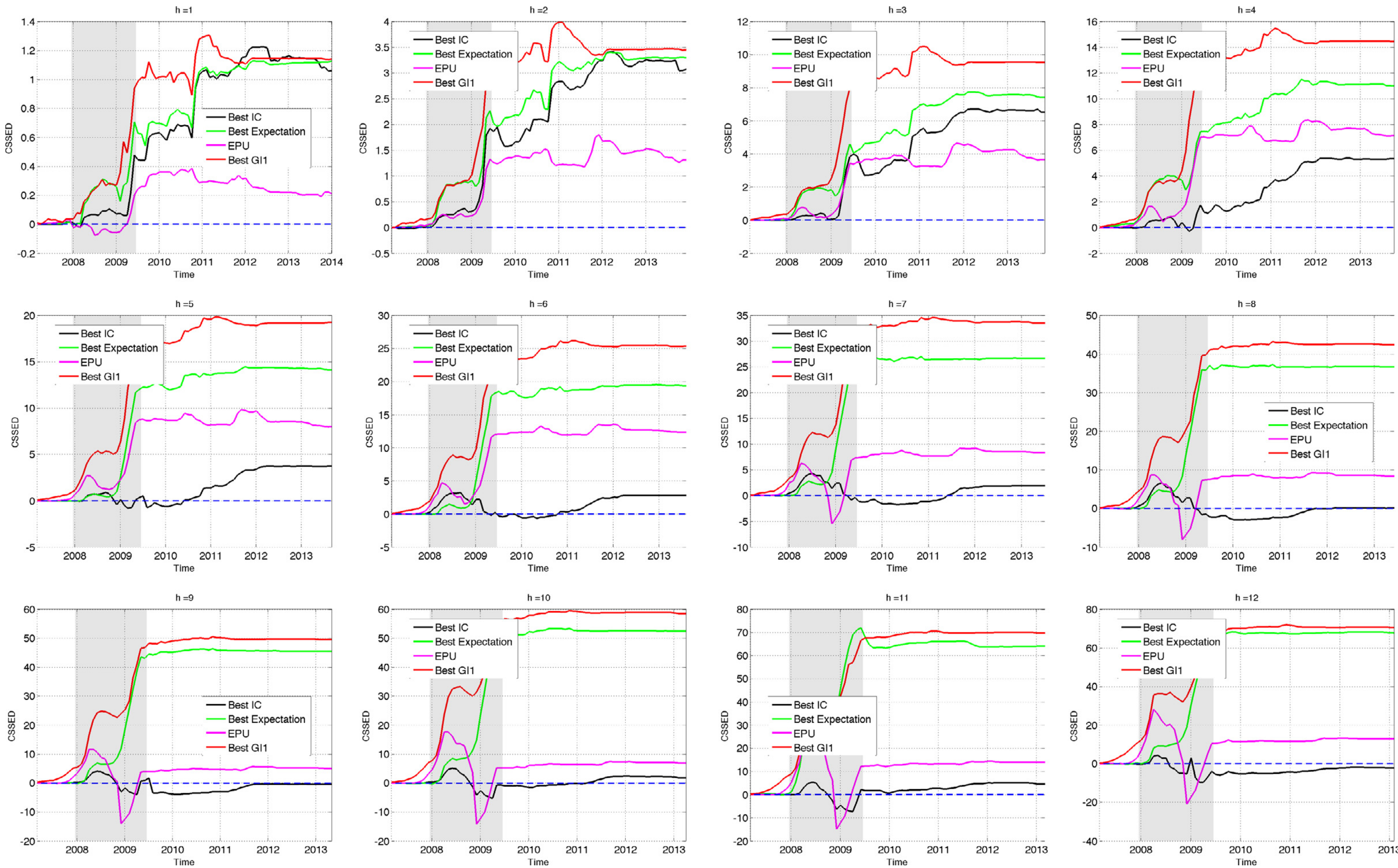

Fig. 3. CSSED comparison ($G1$). *Notes*: For each forecast horizon (h), the graphs show the cumulative sum of squared forecast error differences (CSSED), computed as $CSSED_{m,\tau} = \sum_{\tau=R}^{T}(\hat{e}^2_{bm,\tau} - \hat{e}^2_{m,\tau})$, where $\hat{e}^2_{bm,\tau}$ denotes the squared forecast error of the AR benchmark model and $\hat{e}^2_{m,\tau}$ denotes the alternative model, which may be the best model using IC estimated over the short sample (best IC), the best model using expectations estimated over the short sample (best expectations), or the best model using GI (best GI1). R and T indicate the beginning and end of the forecast evaluation sample, respectively. Values above zero indicate that the alternative model outperforms the benchmark, while values below zero mean that the benchmark outperforms the competing model. Each graph plots the evolution of the CSSED for each forecast horizon.

Table 4
Out-of-sample results for the US unemployment rate in levels (u_t), with alternative keywords; rolling scheme, direct forecasts.

x_{it}	$h=1$	$h=2$	$h=3$	$h=4$	$h=5$	$h=6$	$h=7$	$h=8$	$h=9$	$h=10$	$h=11$	$h=12$
$AR(p)-SS(RMSFE)$	0.206	0.319	0.454	0.561	0.645	0.760	0.877	1.000	1.108	1.209	1.322	1.358
$G1_t$	0.821*	0.765*	0.656*	0.652*	0.644*	0.662*	0.659*	0.665*	0.680*	0.678*	0.673*	0.685*
$G1_{W1,t}$	0.952	0.883	0.967	1.066	1.195	1.276	1.311	1.287	1.252	1.225	1.204	1.231
$G1_{W2,t}$	0.901	0.975	0.967	1.006	1.086	1.051	1.021	1.008	0.971	0.989	0.980	0.907
$G1_{W3,t}$	0.866*	0.802*	0.682*	0.687*	0.704*	0.710*	0.700*	0.682*	0.683*	0.709*	0.731*	0.766*
$G1_{W4,t}$	0.843**	0.777*	0.672*	0.694*	0.691*	0.716**	0.798**	0.798**	0.813*	0.798*	0.793*	0.808*
$G2_t$	1.014	1.064	1.085	1.092	1.062	1.094	1.063	1.062	1.082	1.070	1.040	1.026
$G2_{W1,t}$	1.056++	1.031	1.009	1.005	0.985	0.976	0.985	0.990	0.989	0.998	0.996	0.993
$G2_{W2,t}$	1.066	1.016	1.018	1.025	1.036	1.020	1.039	1.032	1.029	1.025	1.008	1.032
$G2_{W3,t}$	1.006	0.995	1.020+	1.006	0.996	0.993	0.997	0.999	1.000	1.000	0.996	1.001
$G2_{W4,t}$	1.007	1.004	0.995	1.000	1.000	0.993	0.981	0.987	0.994	0.992	0.992	0.992
$G3_t$	1.001	0.987	1.005	0.988	0.942	0.882	0.748	0.659	0.638	0.630	0.627	0.714
$G3_{W1,t}$	1.008	0.959	1.038	1.017	0.906	0.863	0.798	0.814	0.794	0.782	0.801	0.855
$G3_{W2,t}$	1.047	1.036	1.012	0.964	0.909	0.875	0.783	0.743	0.732	0.723	0.740	0.800
$G3_{W3,t}$	1.060	1.023	0.978	0.933	0.942	0.943	0.853	0.753	0.715	0.699	0.696	0.765
$G3_{W4,t}$	1.049	1.184	1.009	1.112	0.858	0.886	0.816	0.757	0.712	0.689	0.693	0.730
$G4_t$	0.976	0.959	0.890	0.886	0.841	0.852	0.851	0.868	0.887	0.899	0.885	0.883
$G4_{W1,t}$	1.014	1.016	0.974	0.981	0.989	1.008	1.011	0.996	0.976	0.975	0.956*	0.954*
$G4_{W2,t}$	1.023	0.960	0.932	0.938	0.968	0.973	0.953	0.912	0.931	0.939	0.928	0.947
$G4_{W3,t}$	1.005	1.004	0.925	0.923	0.903	0.921	0.925	0.934	0.919	0.948	0.941	0.962
$G4_{W4,t}$	1.001	1.000	1.004	1.013	1.020	0.991	1.007	1.004	1.003	1.015	1.017	1.015

Notes: The short sample is 2004.2–2014.2, with the in-sample period being 2004.2–2007.2 and the out-of-sample period being 2007.3–2014.2. $G1_t$ is the Google index for 'jobs', $G2_t$ is the Google index for 'job center', $G3_t$ is the fake Google index for the keyword 'dos' (which is maximally correlated with the unemployment rate in the in-sample period), while $G4_t$ is the Google index for the keywords 'jobs', 'collect unemployment' and 'job center' all combined. For the benchmark model $AR(p)$, the RMSFE is indicated, while for all of the other models the table reports the ratios of the competing model's RMSFE to that of the benchmark at each forecast horizon (h). Thus, a number below 1 indicates that the competitor model beats the $AR(p)$ benchmark, while a number above 1 means that the benchmark performs better. In all panels, *** $(+++)$, ** $(++)$ and * $(+)$ indicate rejection at the 1, 5 and 10% significance levels, respectively, of the Diebold and Mariano (DM, 1995) test for the null of equal forecast accuracy between the $AR(p)$ benchmark estimated over the short sample and the AR model with the alternative leading indicator indicated in the first column when the benchmark is beaten by (beats) the competitor model.

'*dos*' is an acronym for either the US Department Of State or Disk Operating System. We use this alternative web-search indicator (labeled G3 in Table 4) to conduct a horse-race forecast comparison that is identical to the main one, the results of which were presented in Section 4.2. Of course, we do not expect these models to fare well, given that the GI adopted as a leading indicator is completely unrelated to both job searches and unemployment. In fact, looking at Table 4, we can see that models augmented with this false GI are never able to beat the benchmark significantly, with formal DM tests always failing to reject the hypothesis of equal forecast accuracy. This provides further evidence that Google web-search data do have predictive power for the target variable that one wants to forecast if the chosen underlying keyword has a direct link to the phenomenon measured by that variable.[23]

5.3. State-level forecasts

As a final additional robustness check for the predictive properties of the GI, we carry out the standard forecast comparison for each of the 50 states plus DC individually, assessing the percentage of states for which the best model (with the lowest RMSFE) is the one using the G1.[24] Consumers' and employers' expectations are not available at the state level, so the forecast comparison is conducted against the standard $AR(p)$ benchmark and models adopting the initial claims as leading indicators. The descriptive statistics for the monthly unemployment rate, the IC and the GI for each state are reported in the online Appendix (Table B.2, B.3, and B.4), along with graphs showing the dynamics of the G1 and of the unemployment rate for each state (Figures from B.1 to B.5), which tend to be similar for most states.

Before turning to the forecasts, we explore the state-level relationship between Google searches and the unemployment rate further. A simple average of the two variables' contemporaneous correlations across the 51 states is equal to 0.73, confirming the presence of a strong common pattern. Moreover, the correlation is equal to at least 0.75 in 37 of the 51 states (Table B.1 of the online Appendix). The correlations for individual states also remain high at one, two and three lags. We then regress the state-specific correlation on the share of the unemployed that are actually using the internet for their job search, based on the 2011 July Supplement of the Current Population Survey. We find a positive relationship in the contemporaneous correlations that becomes significant at one, two and three lags (Fig. 4; Table B.5 of the online Appendix).

23 We can see from Figure B.14 of the online Appendix that the placebo index works well in the first part of the evaluation sample, given the high correlation with the target variable in the first in-sample period. However, its predictive accuracy deteriorates for the Great Recession at all forecast horizons.

24 In this case, the G1 is equal to the average of 24 raw data downloads, see Section 2.1 for details.

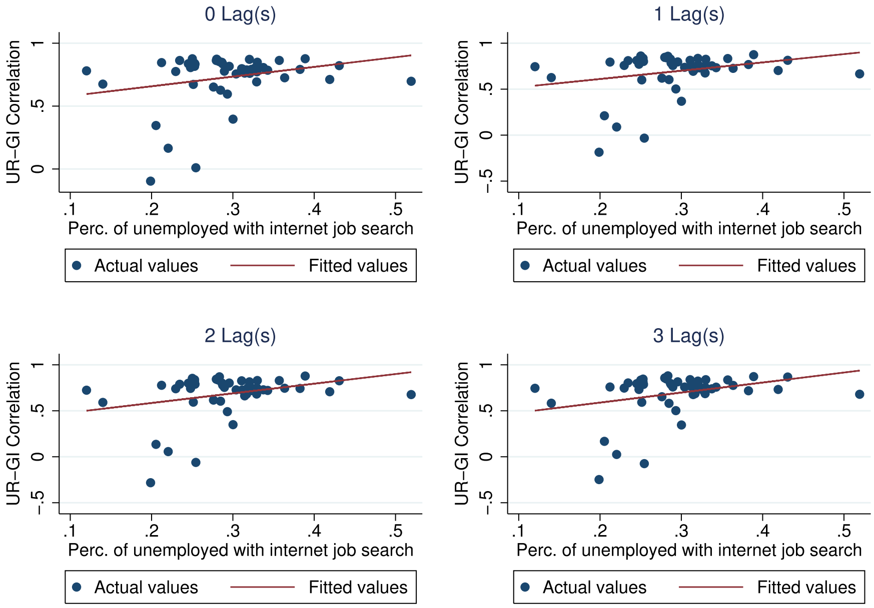

Fig. 4. G1–unemployment rate correlation and the incidence of job-seekers using the internet in each US state. *Notes*: The *x*-axis reports the share of the unemployed that actually use the internet for their job search, as estimated from the 2011 July supplement of the current population survey. The *y*-axis reports the state-specific correlation between the unemployment rate and the G1. Fitted values obtained from the OLS regression (see Table B.5 of the online appendix for regression results).

Moving to the forecast comparisons, for the sake of brevity we will comment here only on the numbers of states in which the models that use the G1 as an exogenous variable provide the most accurate predictions at each forecast horizon.[25]

The Google-based models outperform all of the competing ones in 18 states out of 51 (35% of cases) at one step ahead (Fig. 5); IC-based models perform best in 26 cases (51%); and the $AR(p)$ wins the comparison in seven cases (14%). The performances of Google models improve steadily at longer forecast horizons. At 12 steps ahead, GI models have the lowest RMSFEs in 27 of the 51 states (53%). Thus, the results show that G1-based forecasts are able to improve upon standard ones in a noticeable fraction of US states, even if they perform better overall at the federal level.

Even though a thorough analysis of the unemployment rate predictions at the state level is beyond the scope of this paper, it is important to note here that merely replicating the federal-level exercise confirms the good predictive potential of Google data.

6. Comparison with the Survey of Professional Forecasters

As a final robustness check, we compare the forecasts of our best models with those of the Survey of Professional Forecasters (SPF), a quarterly survey of about 30 professionals conducted by the Federal Reserve Bank of Philadelphia.[26] The professionals receive the survey questionnaire during the first month of each quarter and are asked to return it completed by approximately the middle of the second month of the quarter.[27] Thus, the SPF forecasts include information on labor market data from the first month of each quarter. For the SPF, we consider the median and mean of the individual professionals' forecasts. Conditioning on the same information set, we compare these forecasts with those obtained using a number of different models, chosen from among those with the best forecasting performances. We define these best models as (i) the best Google-based models; (ii) the best non-Google models employing initial claims, one of the four measures of employers' and consumers' expectations used in the rest of the paper, or the economic policy uncertainty index of Baker et al. (2016). For non-Google models, we compare the predictions obtained using models estimated on both the sample starting in 2004 and the longer one starting in 1997. We also include the quarterly forecasts from the two-state steady-state unemployment rate model (SSUR2) of Barnichon and Nekarda (2012), estimated over a fifteen-year rolling window.[28]

We conduct the forecast comparison at four different horizons: a nowcast ($\hat{Q}_{t+0}$), and one-, two- and three-quarter-ahead forecasts ($\hat{Q}_{t+1}$, $\hat{Q}_{t+2}$ and $\hat{Q}_{t+3}$).

Starting with the nowcast, we compute the quarterly forecasts from each model (except for the SSUR2 model, whose forecast is already quarterly) as the average of the realized unemployment rate for the first month and the

25 Full results (including the RMSFE values and the types of models) are reported in Table B.6 of the online Appendix.

26 Available from http://www.phil.frb.org/research-and-data/real-time-center/survey-of-professional-forecasters/.

27 The SPF is issued on around the 15th of February, May, August and November of each year.

28 We would like to thank Regis Barnichon and Chris Nekarda for kindly providing us with the forecasts produced in their paper.

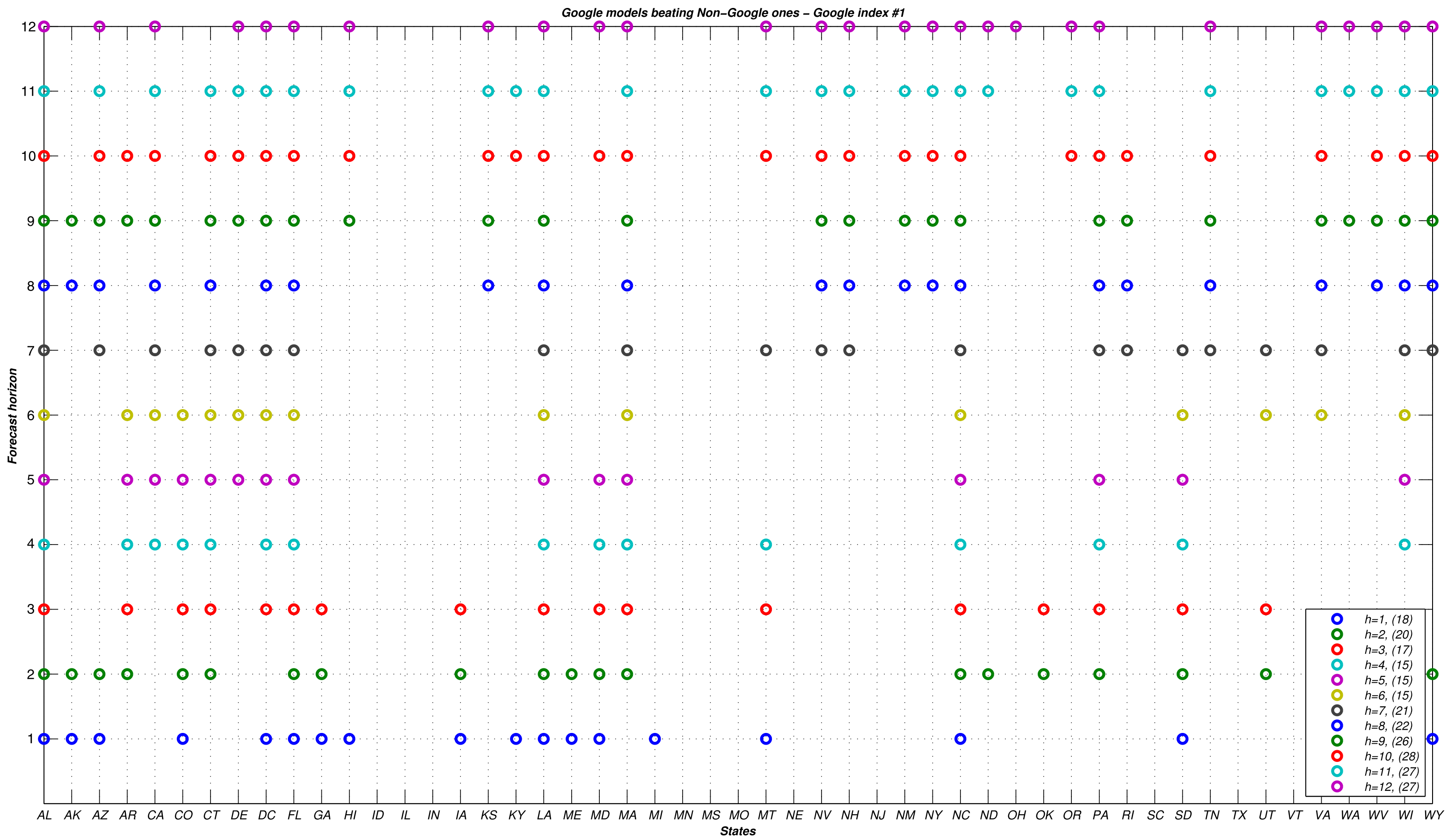

Fig. 5. Numbers of Google models that beat non-Google ones by state. *Notes*: For each of the 50 US states plus DC (on the x-axis), the figure depicts a circle for each of the 12 forecast horizons (on the y-axis) when the best Google model for state j is better than the best non-Google model in terms of RMSFEs. The numbers in parentheses in the legend represent the numbers of states in which the best Google model outperforms the best non-Google model for the various forecast horizons. See Table B.6 of the online appendix for all estimation results.

Table 5
Forecast comparison of the quarterly unemployment rate with the Survey of Professional Forecasters.

x_{it}	$\hat{Q}_{t+0}$		$\hat{Q}_{t+1}$		$\hat{Q}_{t+2}$		$\hat{Q}_{t+3}$	
	RMSFE	Rank	RMSFE	Rank	RMSFE	Rank	RMSFE	Rank
IC_t	0.149	8	0.524	14	0.984*	25	1.269*	21
$IC_{W1,t}$	0.164	18	0.566	23	0.911	22	1.286*	23
$IC_{W2,t}$	0.168	19	0.617	27	0.972	24	1.285*	22
$IC_{W3,t}$	0.151	10	0.573*	25	1.035*	27	1.431*	26
$IC_{W4,t}$	0.141	3	0.548	21	1.032*	26	1.433*	27
EEM_t	0.142	5	0.536*	18	0.938	23	1.303	24
$EENM_t$	0.139	2	0.422	4	0.711	5	1.004	6
CE_t	0.181**	24	0.500	8	0.759*	8	0.990	5
CE_t^{6M}	0.190*	26	0.533	17	0.732	6	0.944	4
EPU_t	0.168**	20	0.523*	13	0.882*	20	1.385	25
$G1_t$	**0.138**	1	0.404	3	**0.598**	1	**0.781**	1
$G1_{W1,t}$	0.162	17	0.739	28	1.348	28	1.717	28
$G1_{W2,t}$	0.188**	25	0.608	26	0.886	21	1.221	20
$G1_{W3,t}$	0.147	7	0.393	2	0.639	2	0.881*	3
$G1_{W4,t}$	0.155*	13	0.439	5	0.656	3	0.870	2
$SSUR2$	0.152	11	**0.377**	1	0.671	4	1.008	7
SPF^{mean}	0.207**	28	0.470	6	0.769	9	1.114	15
SPF^{median}	0.199**	27	0.474	7	0.754	7	1.109	11
$IC_t - LS$	0.149	9	0.513	12	0.799*	14	1.099*	10
$IC_{W1,t} - LS$	0.157	16	0.529	16	0.813	16	1.113*	14
$IC_{W2,t} - LS$	0.155	14	0.529	15	0.809	15	1.112*	13
$IC_{W3,t} - LS$	0.146	6	0.509*	11	0.793*	13	1.098*	9
$IC_{W4,t} - LS$	0.141	4	0.501	9	0.784*	11	1.087*	8
$EEM_t - LS$	0.156	15	0.541*	19	0.824	17	1.119	16
$EENM_t - LS$	0.153	12	0.507	10	0.785	12	1.129	17
$CE_t - LS$	0.177**	22	0.565	22	0.842*	19	1.166	19
$CE_t^{6M} - LS$	0.174*	21	0.543	20	0.773	10	1.112	12
$EPU_t - LS$	0.181**	23	0.567*	24	0.840*	18	1.152	18

Notes: The table shows the quarterly forecast comparison of our model forecasts with the SPF median and mean forecasts. Q_{t+0} denotes the current quarter. The forecast origin is the end of the second week of the second month of each quarter (e.g., May 15, 2008, for the nowcast of the 2008Q2 unemployment rate). The out-of-sample period is 2007.Q2–2014.Q1. SPF^{median} and SPF^{mean} are the median and the mean of the SPF forecasts, respectively. *SSUR*2 is the steady-state unemployment rate model with two states, proposed by Barnichon and Nekarda (2012) and estimated over a fifteen-year rolling window. For each model, the quarterly forecasts are computed as the average of the unemployment rate for the first month of the quarter and the one- and two-month-ahead forecasts computed right before the 15th of the second month of each quarter. We present the results for the forecasts with all of the leading indicators considered in this paper for the short sample, along with those for the long sample (LS). For each model we present the root MSE (RMSFE), the rank, and the significance of the DM test. We indicate the benchmark model with the lowest RMSFE for each forecast horizon in boldface.
***, ** and * indicate rejection of the null of equal forecast accuracy for the DM test at the 1, 5, and 10% significance levels, respectively.

1- and 2-month-ahead forecasts generated in the middle of the second month of the quarter. For quarterly forecasts at longer horizons, we just aggregate the corresponding monthly forecasts (three to five steps ahead for $\hat{Q}_{t+1}$, six to eight for $\hat{Q}_{t+2}$, and nine to eleven for $\hat{Q}_{t+3}$).

The left side of Table 5 reports the RMSFE from each set of best models and the median of the SPF forecasts over the period 2007Q2–2014Q1, along with the DM test where the benchmark is the best model; that is, the model with the lowest RMSFE (in boldface). The Google-based forecast beats the SPF forecast at every forecast horizon. The reduction in RMSFE is the greatest (−44%) when nowcasting the unemployment rate ($\hat{Q}_{t+0}$), with the DM rejecting the null of equal forecast accuracy. The advantage of the GI model diminishes to 17% at $\hat{Q}_{t+1}$, increasing again to 26% and 46% respectively at $\hat{Q}_{t+2}$ and $\hat{Q}_{t+3}$. Our best Google model also provides a smaller RMSFE than that of any other estimated model for three out of four forecast horizons. Only at $(t+1)$ does a 'state-of-the-art' model such as the SSUR2 of Barnichon and Nekarda (2012), based on labor force flows, outperform the GI model, with a RMSFE that is 7% lower.

Figure B.6 in the online Appendix displays the nowcasting forecast errors from the best models of Table 5, in addition to the SPF forecast. It is clear from the figure that the model that includes the GI performs the best in most periods, and in particular when the Great Recession worsened in 2008Q4. The Google model tends to achieve forecast errors that are closer to zero, whereas the median of the SPF tends to under-predict the real unemployment rate. The same happens to the SSUR2 model, which seems to over-predict only slightly at the end of the Great Recession.

Thus, simple linear AR models augmented with the GI as a leading indicator outperform the predictions of the professional forecasters' models at turning points following unprecedented economic contractions, a time when it is very difficult to make accurate predictions. In fact, Figure B.7 of the online Appendix shows that the dispersion of the professional forecasters' predictions was very high during the Great Recession, and all of their forecasts tended to under-predict the real quarterly unemployment rate. Moreover, the low RMSFEs of Google-based models are not due only to the good performance during the deepening of the Great Recession, but are confirmed over the whole out-of-sample interval, as is also shown by the dynamics over time of the CSSED reported in Figure B.8 of the online Appendix.

7. Conclusions

This paper assesses the relevance of a Google index (GI), based on internet job searches performed through Google, for predicting the US monthly unemployment rate. Google-based models performed particularly well around the turning point at the beginning of the Great Recession, with their relative performance stabilizing thereafter.

Popular time series specifications that are augmented with this indicator improve their out-of-sample forecasting performances for forecast horizons of one to twelve months. Google-based models also generally fare better than other similar models estimated on the same or longer time spans and using the initial claims (IC), employers' or consumers' expectations or the economic policy uncertainty index of Baker et al. (2016) as a leading indicator. At one and two months ahead, models using employers' expectations as an exogenous variable can improve on Google ones only when using a longer time series, while at three to twelve months ahead, web-search-based models always do better. We explore the determinants of the good predictive performances of Google-based models further by looking at labor market transitions at one- and twelve-month intervals separately for individuals who did and did not search for a job online. We find that employed individuals who search for a job online are more likely to be unemployed twelve months later. A higher incidence of unemployment among individuals who are looking for jobs online is also found for those who were unemployed or inactive a year earlier. These results suggest that the good forecasting performance of the Google-based indicator, which improves with the forecast horizon, can be attributed to the fact that individuals start to look for employment online long before losing their job or being reclassified as unemployed instead of inactive.

The results of the forecast comparison at the state level are less clear cut than those at the federal level, but still point to the substantive forecasting power of Google-based models. Moreover, we find that the correlation between the unemployment rate and the Google index is significantly stronger when a greater percentage of the unemployed are using the internet for their job search.

Finally, conditioning on the same information set, the best Google-augmented predictions also outperform the forecasts from both the Survey of Professional Forecasters, conducted by the Philadelphia Fed, and state-of-the-art models based on labor flows.

In summary, using ten years of available data, we find ample convincing evidence of the usefulness of indices of online job searches for forecasting US unemployment, especially at turning points and for medium to long forecast horizons. Should these results be confirmed as more data become available over time, we believe that web-based leading indicators would have to be included routinely in models for predicting unemployment.

Acknowledgments

The views expressed in this paper are those of the authors and do not necessarily reflect those of the Bank of Italy. We would like to thank the editor Michael McCracken, an anonymous associate editor and an anonymous referee for their helpful comments that helped us greatly in improving the previous version of the paper. We would also like to thank F. Busetti, D. Depalo, G. Elliott, D. Giannone, G. Gonzalez-Rivera, O. Jorda, I. Komunjer, T.-H. Lee, F. Lotti, G. Mesters, R. Mosconi, C. Perna, T. Proietti, A. Rosolia, B. Rossi, P. Sestito, H. Varian, A. Timmermann, A. Ullah, J. Wright and K. Zimmermann for their useful suggestions. We also thank seminar participants at UCSD and UCR Economics Department, UPF Department of Economics and Business, Tor Vergata University, IMT, the 32nd International Symposium on Forecasting, the 2nd International Conference in Memory of Carlo Giannini, the XVIII SNDE Symposium, the 45th Scientific Meeting of the Italian Statistical Society, the Fourth Italian Congress of Econometrics and Empirical Economics, Fondazione ENI Enrico Mattei and LUISS 'Guido Carli' University for their useful comments. Last but not least, we would like to thank R. Barnichon and C. J. Nekarda for kindly sharing their forecasts.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at http://dx.doi.org/10.1016/j.ijforecast.2017.03.004.

References

Antenucci, D., Cafarella, M., Levenstein, M.C., Re, C., & Shapiro, M.D. (2014). *Using social media to measure labor market flows*. NBER Working Papers, 20010.

Artola, C., & Galan, E. (2012). *Tracking the future on the web: construction of leading indicators using Internet searches*. Documentos Ocasionales (1203), Bank of Spain.

Askitas, N., & Zimmermann, K. F. (2009). Google econometrics and unemployment forecasting. *Applied Economics Quarterly*, *55*, 107–120.

Askitas, N., & Zimmermann, K. F. (2015). The Internet as a data source for advancement in social sciences. *International Journal of Manpower*, *36*(1), 2–12.

Baker, S., Bloom, N., & Davis, S. J. (2016). Measuring economic policy uncertainty. *The Quarterly Journal of Economics*, *131*(4), 1593–1636.

Baker, S., & Fradkin, A. (2014). The impact of unemployment insurance on job search: evidence from Google search data. *Review of Economics and Statistics*, (forthcoming).

Barnichon, R., & Nekarda, C. J. (2012). The ins and outs of forecasting unemployment: using labor force flows to forecast the labor market. *Brookings Papers on Economic Activity*, 83–131.

Billari, F., D'Amuri, F., & Marcucci, J. (2016). *Forecasting births using Google*. Bank of Italy Discussion Paper (forthcoming).

Choi, H., & Varian, H. (2012). Predicting the present with Google Trends. *Economic Record*, *88*, 2–9.

Da, Z., Engelberg, J., & Pengjie, G. (2011). In search of attention. *Journal of Finance*, *5*, 1461–1499.

D'Amuri, F. (2009). *Predicting unemployment in short samples with Internet job search query data*. MPRA working paper No. 18403.

D'Amuri, F., & Marcucci, J. (2009). *'Google it!' Forecasting the US unemployment rate with a Google job search index*. ISER Working paper series, 2009-32.

DeLong, J. B., & Summers, L. H. (1986). Are business cycles symmetrical? In R. J. Gorton (Ed.), *The American business cycle, continuity and changes*. Chicago: University of Chicago Press for NBER.

Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business and Economic Statistics*, *13*, 253–263.

East, C. N., & Kuka, E. (2015). Reexamining the consumption smoothing benefits of unemployment insurance. *Journal of Public Economics*, *132*, 32–50.
Edelman, B. (2012). Using Internet data for economic research. *The Journal of Economic Perspectives*, *26*(2), 189–206.
Einav, L., & Levin, J. (2013). Economics in the age of big data. *Science*, *346*(6210), 715–721.
Ettredge, M., Gerdes, J., & Karuga, G. (2005). Using web-based search data to predict macroeconomic statistics. *Communications of the ACM*, *48*(11), 87–92.
Fondeur, Y., & Karamé, F. (2013). Can Google data help predict French youth unemployment? *Economic Modelling*, *30*, 117–125.
Ginsberg, J., Mohebbi, M. H., Patel, R. S., Brammer, L., Smolinski, M. S., & Brilliant, L. (2009). Detecting influenza epidemics using search engine query data. *Nature*, *457*, 1012–1014.
Golan, A., & Perloff, J. M. (2004). Superior forecasts of the U.S. unemployment rate using a nonparametric method. *The Review of Economics and Statistics*, *86*(1), 433–438.
Hastie, T. J., & Tibshirani, R. J. (1990). *Generalized additive models*. London: Chapman and Hall Ltd..
Jurado, K., Ludvigson, S. C., & Ng, S. (2015). Measuring uncertainty. *The American Economic Review*, *105*(3), 1177–1216.
King, M., Ruggles, S., Alexander, J.T., Flood, S., Genadek, K., & Schroeder, M.B. et al. (2010). *Integrated public use microdata series, current population survey: Version 3.0*. [Machine-readable database]. Minneapolis, MN: Minnesota Population Center [producer and distributor].
Koop, G., & Potter, S. M. (1999). Dynamic asymmetries in US unemployment. *Journal of Business and Economic Statistics*, *17*(3), 298–312.
McLaren, N., & Shanbhorge, R. (2011). Using Internet data as economic indicators. *Bank of England Quarterly Bulletin, Second Quarter*,.
McQueen, G., & Thorley, S. (1993). Asymmetric business cycle turning points. *Journal of Monetary Economics*, *31*, 341–362.
Mohebbi, M., Vanderkam, D., Kodysh, J., Schonberger, R., Choi, H., & Kumar, S. (2011). Google correlate whitepaper. Google, www.google.com/trends/correlate/whitepaper.pdf.
Montgomery, A. L., Zarnowitz, V., Tsay, R. S., & Tiao, G. C. (1998). Forecasting the US unemployment rate. *Journal of the American Statistical Association*, *93*(442), 478–493.
Neftci, S. N. (1984). Are economic time series asymmetric over the business cycles? *Journal of Political Economy*, *85*, 281–291.
Orlik, A., & Veldkamp, L. (2014). *Understanding Uncertainty Shock and the Role of Black Swans*, NBER Working paper 20445, National Bureau of Economic Research. https://ideas.repec.org/p/nbr/nberwo/20445.html.
Preis, T., Moat, H. S., & Stanley, E. (2013). Quantifying trading behavior in financial markets using Google Trends. *Nature Scientific Reports*, Article 1684.
Proietti, T. (2003). Forecasting the US unemployment rate. *Computational Statistics and Data Analysis*, *42*, 451–476.
Rothman, P. (1998). Forecasting asymmetric unemployment rates. *The Review of Economics and Statistics*, *80*(1), 164–168.
Sichel, D. E. (1993). Business cycle asymmetry: a deeper look. *Economic Inquiry*, *31*, 224–236.
Stock, J. H., & Watson, M. W. (2003). Forecasting output and inflation: the role of asset prices. *Journal of Economic Literature*, *41*(3), 788–829.
Suhoy, T. (2009). *Query indices and a 2008 downturn*. Bank of Israel Discussion Paper (2009.06).
Troy, G., Perera, D., & Sunner, D. (2012). Electronic indicators of economic activity. *Australian Reserve Bank, Economic Bulletin*, June 1–12.
Tuhkuri, J. (2016). *Forecasting unemployment with Google searches*. ETLA Working Paper No. 35.
Vosen, S., & Schmidt, T. (2011). Forecasting private consumption: survey-based indicators vs. Google trends. *Journal of Forecasting*, *30*, 565–578.
Welch, I., & Goyal, A. (2008). A comprehensive look at the empirical performance of equity premium prediction. *Review of Financial Studies*, *21*(4), 1455–1508.

Francesco D'Amuri works as an advisor at the Economics, Statistics and Research Department of the Bank of Italy. He has also been a research affiliate with IZA—Bonn since 2011. His main interests are applied labor and public economics, with a particular focus on migration, industrial relations and short term economic forecasting. He has been a visiting scholar at the University of California Davis, the Federal Reserve Bank of San Francisco, and the University of California Los Angeles. His work has appeared in the *Journal of the European Economic Association* and the *European Economic Review*. He holds a B.Sc. in Economics from Bocconi (Italy) University and a Ph.D. in Economics from the University of Essex (UK).

Juri Marcucci works as an advisor at the Economics, Statistics and Research Department of the Bank of Italy. His research interests are in volatility modeling and forecasting, big data, forecast evaluation, forecasting with nonlinear time series, the predictability of returns, nonlinear panel data models, and credit risk cyclicality. He has been a lecturer at the University of Bologna and the University of Rome Tor Vergata. He has also been a visiting scholar at the University of California San Diego, the Federal Reserve Bank of Boston and UPF. His work has appeared *in the Journal of Econometrics*, *International Journal of Forecasting, Studies in Nonlinear Dynamics and Econometrics*, and the *Journal of Banking and Finance*. He holds a Ph.D. in Economics from the University of California San Diego.