

D'Amuri, Francesco; Marcucci, Juri

Working Paper

'Google it!' Forecasting the US unemployment rate with a Google job search index

ISER Working Paper Series, No. 2009-32

Provided in Cooperation with:

Institute for Social and Economic Research (ISER), University of Essex

Suggested Citation: D'Amuri, Francesco; Marcucci, Juri (2009) : 'Google it!' Forecasting the US unemployment rate with a Google job search index, ISER Working Paper Series, No. 2009-32, University of Essex, Institute for Social and Economic Research (ISER), Colchester

This Version is available at:

<https://hdl.handle.net/10419/92170>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

“Google it!”

Forecasting the US unemployment rate with a Google job search index

Francesco D’Amuri

Bank of Italy
ISER, University of Essex

Juri Marcucci

Bank of Italy

No. 2009-32
November 2009



INSTITUTE FOR SOCIAL
& ECONOMIC RESEARCH

“Google it!”

Forecasting the US unemployment rate with a Google job search index

Non-technical summary

A recent article published on *Nature* developed and tested a simple model to forecast physician visits due to influenza-like illness (ILI) in a particular region using as a single explanatory variable the ILI-related query fraction on total queries as recorded by Google search engine data.

Following the high popularity of the internet as a job-search mean, and the increasing need of up-to-date economic indicators in the current economic downturn, this approach has recently been extended to unemployment forecasting. In particular, the Google Index (GI), equal to the incidence of google job-search-related queries over total queries, proved to have predictive power in forecasting unemployment developments in Germany and Israel; another study used it to predict the initial unemployment claims in the US.

The main limitation of the explanatory variable based on Google data is that it could be partly driven by on-the-job search, rather than unemployed job search activities which are the focus of this paper. Another limitation is due to the fact that not all workers have access to the internet, and it is also presumable that workers using the internet for a job search are not randomly selected among job-seekers. This should be a minor issue, given the increasing popularity of the internet as a job search method and also because a bias in the estimates would emerge only if shocks hit the unemployed using/not using the internet for job search in a different way.

In this article, we test the relevance of a Google job-search index as a leading indicator for unemployment dynamics in the United States, finding that it dramatically increases the precision of the forecasts.

In particular we find that, models including this index among the explanatory variables outperform more than 300 alternative popular time series specifications (both linear and non linear) when forecasting at one, two and three steps ahead. The indicator has power not only at the federal but also at the state level: for 70 per cent of the states the best model in terms of forecast accuracy includes the GI. Models augmented with this Index also outperform the estimates released in the Survey of Professional Forecasters conducted by the Federal Reserve of Philadelphia, having a mean squared error lower by an order of magnitude.

“Google it!”

Forecasting the US unemployment rate with a Google job search index

Francesco D’Amuri
Bank of Italy
and
ISER, University of Essex

Juri Marcucci*
Bank of Italy

November 13, 2009

Abstract

We suggest the use of an Internet job-search indicator (the Google Index, GI) as the best leading indicator to predict the US unemployment rate. We perform a deep out-of-sample forecasting comparison analyzing many models that adopt both our preferred leading indicator (GI), the more standard initial claims or combinations of both. We find that models augmented with the GI outperform the traditional ones in predicting the monthly unemployment rate, even in most state-level forecasts and in comparison with the Survey of Professional Forecasters.

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

JEL Classification: C22, C53, E27, E37, J60, J64.

*The views expressed are those of the authors and do not necessarily reflect those of the Bank of Italy. We wish to thank Fabio Buseti, Oscar Jorda, Francesca Lotti, Alfonso Rosolia and Paolo Sestito for their useful comments. Francesco D’Amuri gratefully acknowledges support from the Economic and Social Research Council. Emails: francesco.damuri@gmail.com (Francesco D’Amuri), juri.marcucci@gmail.com (Juri Marcucci, corresponding author). Correspondence address: Bank of Italy, Economic Research Department, Via Nazionale 91, 00184, Rome, Italy.

1 Introduction

Quantitative data on internet use are becoming quickly available and will constitute an invaluable source for economic analysis in the near future. Following the growing popularity of the internet as a job search tool and the increasing need of reliable and updated unemployment forecasts, especially in the recent economic downturn, in this article we suggest the use of the Google index (GI) as the best leading indicator to predict the US unemployment rate.¹ We test the predictive power of this new leading indicator based on Google job-search-related query data by means of a deep out-of-sample comparison among more than five hundred forecasting models which differ along three dimensions: i) the exogenous variables adopted as leading indicators, ii) the econometric specification, and iii) the length of the estimation sample. In particular, we estimate standard time series (ARMA) models and we augment them with the initial claims (the IC, a widely accepted leading indicator for the US unemployment rate), the GI, or combinations of both. In carrying out our comparison, we include both linear and non-linear models, because the former typically capture short-run developments, while the latter can better approximate the dynamics of the unemployment rate during economic contractions. In our forecasting horse-race, we compare models estimated over samples of different length, because the GI is only available since 2004, while the IC are available since 1967. Indeed, an exercise comparing the forecasting performance of models estimated on the short sample only (starting in 2004) would be of little practical relevance if models estimated on the longer sample (starting in 1967) were better at predicting the unemployment rate.

We find that models augmented with the GI significantly outperform the more traditional ones in predicting the US unemployment rate: when forecasting at one-step ahead

¹The time series of US unemployment rate is certainly one of the most studied in the literature. Proietti (2003) defines this series as the ‘testbed’ or the ‘case study’ for many (if not most) non-linear time series models. In fact, many papers have documented its asymmetric behavior. Neftci (1984), DeLong and Summers (1986) and Rothman (1998) document the type of asymmetry called *steepness* for which unemployment rates rise faster than they decrease. Sichel (1993) finds evidence for another type of asymmetry called *deepness* in which contractions are deeper than expansions. McQueen and Thorley (1993) find *sharpness* for which peaks tend to be sharp while troughs are usually more rounded.

the mean squared error (MSE) of our best model using GI as a leading indicator (0.0166) is 29% lower than the best model not including it, regardless of the estimation sample and the econometric specification. Relative forecast accuracy increases at longer forecast horizons: at three steps ahead, when using the GI the MSE decreases by 40%.

As a robustness check, we test the predictive power of the GI estimating the same set of models on the most commonly used transformations for the time series of the unemployment rate,² finding similar results. As a further check, we forecast the unemployment rate in each of the 51 US states (including District of Columbia) with the same set of models, finding that in more than 70% of the cases, models including the GI outperform all the others. Finally, we construct a group of quarterly forecasts of the unemployment rate using 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. Even in this case we find that models using the GI outperform the professionals' forecasts, showing a lower MSE by an order of magnitude.

Furthermore, we select the best models in terms of the lowest MSE and we test both for equal forecast accuracy and forecast encompassing to assess their out-of-sample forecast ability. We also test our best models in terms of their superior predictive ability, which allows us to control for the effects of data-snooping biases. To do this we employ the Reality Check test suggested by White (2000).

The first article using Google data (Ginsberg et al., 2009) estimates the weekly '*influenza*' activity in the US using an index of the health seeking behavior equal to the incidence of *influenza*-related internet queries. To the best of our knowledge, this is the first paper using this kind of internet indicator to forecast the unemployment rate in the US. However, there have already been some works for other countries, in particular for Germany (Askatas and Zimmermann, 2009), Italy (D'Amuri, 2009) and Israel (Suhoy, 2009), while Choi and Varian (2009) use the GI to predict initial unemployment claims

²In particular, we use the following transformations: logit (as in Koop and Potter, 1999 or Wallis, 1987), first differences (as in Montgomery et al, 1998), logarithm, log-linear detrended or HP-filtered in logs (as in Rothman, 1998).

for the US. Based on our results for the unemployment rate, we believe that there will be more and more applications using Google query data in the future also in other fields of economics.

The paper is organized as follows: in Section 2 we describe the data used to predict the US unemployment rate, with a particular emphasis on the Google index. In Section 3 we discuss the models employed to predict the US unemployment rate, while in Section 4 we compare the out-of-sample performance of such models. Finally, in Section 5 we perform some robustness tests, checking the predictive ability of models augmented with the GI at the state level, comparing the results of the federal estimates both with the quarterly estimates of the SPF and some nonlinear models typically deemed as the best forecasting models in the literature. In section 6 we present our conclusions.

2 Data

The data used in this paper come from different sources. The seasonally adjusted monthly unemployment rate is the one released by the Bureau of Labor Statistics and comes from the Current Employment Statistics and the Local Area Unemployment Statistics for the national and the state level, respectively. Unemployment rates for month t refer to individuals who do not have a job, but are available for work, in the week including the 12th day of month t and who have looked for a job in the prior 4 weeks ending with the reference week. For the federal level the available sample is 1948.1-2009.6, while for the state level the data on unemployment are available from 1976.1 to 2009.6. We complement these data with the weekly seasonally adjusted Initial Claims (IC) released by the U.S. Department of Labor³, a well-known leading indicator for the unemployment rate (see for example Montgomery et al. 1998). The weekly IC for the US are available from 1967.1 until 2009.6, while for the single states they are only available from 1986.12.

The exogenous variable specific to this study is the weekly Google Index (GI) which

³Since seasonally adjusted data are issued only at the national level, we have performed our own seasonal adjustment for the state-level data using Tramo-Seats.

summarizes the job searches performed through the Google website. The data are available almost in real time starting with the week ending on January 10, 2004 and report the incidence of queries using the keyword “*jobs*” on total queries performed through Google in the relevant week.⁴ The values of the index, available free of charge,⁵ are normalized with a value equal to 100 for the week with the highest incidence.

We chose to use the keyword “*jobs*” as an indicator of job search activities for two reasons. First, since absolute search volumes are not available, we identify the most popular keywords looking at relative incidences. In these terms, we found that the keyword “*jobs*” was the one showing the highest incidence among different job-search-related keywords. Even if we do not know the absolute search volumes, we can compare the relative incidences of searches for the keyword “*jobs*” with other extremely popular keywords searches. In particular, in Figure 1, we plot the monthly averages for the values of the GI for the keywords: “*facebook*”, “*youtube*”, “*jobs*” and “*job offer*”. We notice that, when the incidence of keyword searches for “*facebook*” was at its highest level in the interval considered here, the GI was slightly below the value of 80, while the GI for the keyword “*jobs*” was slightly above 20. This means that in that period there was more than one keyword search for “*jobs*” for each four searches for “*facebook*”. The results are similar when conducting the comparison with the keyword “*youtube*”, another popular search. Finally, the alternative job-search-related keyword “*job offers*” reaches very low values of the GI (basically zero) in the interval. Apart from its popularity, the second reason why we chose the keyword “*jobs*” is that we believe that it is widely used across the broadest range of job seekers. We could have augmented it with other job-search-related keywords, such as “*unemployment benefits*” or “*state jobs*”. This would have increased the volume of searches underlying the value of the GI. But, at the same time, the information conveyed by these keywords is related to particular subgroups of the population, and the presence of demand or supply shocks specific to these subgroups could bias the values of the GI and its ability to predict the overall unemployment rate.

⁴We have adjusted both the weekly and the monthly indicators for seasonality using Tramo-Seats.

⁵www.google.com/insights/search/#. The data used in this article were downloaded on July 29, 2009.

However, the variable has its limitations: individuals looking for a job through the internet (jobs available through the internet) may well be not randomly selected 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's job search is believed to follow the anti-cyclical variation of job separation rates, on-the-job search is normally assumed to be cyclical. We acknowledge that this can induce some bias in our preferred leading indicator the GI.

In the empirical analysis we align the GI and IC data with the relevant weeks for the unemployment survey. In other words, when constructing the GI or the IC for month t , we take into consideration the week including the 12th of the month and the three preceding weeks, exactly the same interval used to calculate the unemployment rate for month t reported in official statistics. 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 in order to calculate the unemployment rate (see Figure 2 for a visual description of the alignment procedure).

Table 1 reports the descriptive statistics for various transformations of the US unemployment rate and both leading indicators (IC and the GI, both weekly and monthly). In the Appendix we also show the descriptive statistics of the IC and the GI both for the United States as a whole and for each single state (Tables A.1, A.2 and A.3). The IC for the US are publicly available through the Department of Labor website starting with January 1967, while those for the single states are available since December 1986. The monthly averages of the IC have almost always right-skewed distributions and are highly non-normal (we always reject the null of normality with the Jarque-Bera test). The monthly averages of the GI (which starts in January 2004) are also right-skewed with non-normal distribution, except for Alaska and Maine. The weekly IC and GI (those with the subscript $wj, j = 1, \dots, 4$) show similar features. From Table A.4 in the Appendix we can infer that the unemployment rate also has a right-skewed distribution and a high kur-

tosis which make the series non-normal as suggested by the Jarque-Bera test that almost always rejects the null hypothesis of normality. The same happens for the unemployment rate of each single state except for Colorado.

In Figure 3 and 4, we plot separately the national unemployment rate and our exogenous variables adopted as leading indicators over the relevant sample periods. In Figure 3, we plot the unemployment rate and the IC over the sample period 1967:1-2009:6, according to the availability of IC. Figure 4 depicts instead the unemployment rate along with the IC as well as the Google ‘job’ search index over the sample 2004:1-2009:6. These latter indexes are rescaled with respect to the maximum value of each series over the sample. In both cases the two series show similar patterns, with both IC and the GI seeming to be leading indicators for the unemployment rate. This behavior is confirmed by the correlations: focusing on the 2004:1-2009:6 period, we can see that both the GI and the IC are highly correlated with the level and with the first differences of the unemployment rate (see Table 2). In particular, the correlations of the GI with the first differences are higher than those of the IC, suggesting that this alternative indicator can be rather helpful for predicting not only the level of the unemployment rate but also its changes.

Before proceeding with our forecasting exercise and the in-sample estimation of our set of models, we have checked for non-stationarity of the US unemployment rate by computing a robust univariate unit root test for the integration of the series. We have performed the Augmented Dickey-Fuller test with GLS de-trending (ADF-GLS) suggested by Elliott et al. (1996). This test is similar to the more standard Dickey-Fuller t test but it applies GLS de-trending before the series is tested with the ADF test. Compared with the standard ADF test, ADF-GLS test has the best overall performance in terms of small-sample size and power. Table 3 reports the results of this unit root test both considering a constant (superscript μ) and a constant and trend (superscript τ) as exogenous regressors. We run these tests both for the full sample, i.e. 1967.1-2009.6, and for the short sample, i.e. 2004.1-2009.6. We report the unit root test results for the unemployment rate in

levels u_t , and for other transformations typically used in the literature on the US.⁶

Looking at u_t , the ADF-GLS ^{μ} test fails to reject the null of a unit root for the full sample, but strongly rejects (at 1%) the null for the short sample. Similarly, the ADF-GLS ^{τ} test fails to reject the null of a unit root on the full sample but it does reject the null on the short sample, indicating that the series of unemployment is stationary over this shorter sample. For all the other transformations, the ADF-GLS tests suggest an overall rejection of the null of a unit root only when the null is non-stationarity around the mean over the short sample. The test fails to reject over the full sample, except for the transformation u_t^{LHP} . We should also notice that over the short sample the ADF-GLS ^{τ} tests are very close to the 10% critical value.

However, in the literature most works impose the presence of a unit root using the first differences of the unemployment rate for forecasting purposes. For example, Montgomery et al. (1998) argue that unit-root non-stationarity might be hard to justify for the US unemployment rate series, but nevertheless adopt an ARIMA(1,1,0)(4,0,4) as their benchmark model for short-term forecasting. In what follows, we adopt a more general approach modeling both the level and the first differences of the unemployment rate series because we are interested in finding the best model for short-term forecasting and not in modeling the long-term dynamics of the series.

3 Forecasting models

In our forecasting exercise we compare a total of 520 linear ARMA models for the variable $u_t - u_{t-1}$, which denotes the first differences of the US unemployment rate. As a robustness check, we also estimate the same set of models on the level and the most commonly used transformations for u_t : logarithm, logit, first differences, log-linear detrended or HP-filtered in logs. For the sake of brevity, and since all main results are confirmed when

⁶We use in particular, the log-level ($\log(u_t)$), the logistic transformation ($u_t^{logit} = \log(\frac{u_t}{1-u_t})$) suggested by Koop and Potter (1999) following Wallis (1987) to make the series unbounded, the log-linear de-trended ($u_t^{LTD} = \log(u_t) - \hat{a} - \hat{b}t$) and the HP-filtered series in log (u_t^{LHP}) both suggested by Rothman (1998).

using these transformations, we will only comment on the estimates obtained from the first differences of the unemployment rate. A full list of the models estimated on this series and their forecasting performance can be found in Table A.5 of the Appendix.

We estimate 384 AR, ARMA and ARMAX models that can be grouped in three broad categories:

- a) models *not including* the GI as an exogenous variable and estimated on the full sample (in sample 1967:1-2007:2; out of sample 2007:3-2009.6)
- b) models *not including* the GI as an exogenous variable but estimated on the short sample, for which Google data are available (in sample 2004:1-2007:2; out of sample 2007:3-2009.6)
- c) models *including* the GI as an exogenous variable and estimated on the short sample (in sample 2004:1-2007:2; out of sample 2007:3-2009.6).

Within these three broad groups we estimate exactly the same set of models, both in terms of lag specification and of exogenous variables included, with the GI indicator added as an additional independent variable in the last, otherwise identical, set of models.

We also estimate, on the short sample, an additional set of 136 models including different combinations of lag structures and exogenous variables. The rationale of repeating our forecasting exercise along three dimensions is straightforward. The inclusion of the GI among the exogenous variables limits the length of the estimation interval, given that the indicator is available only since 2004.1. An exercise comparing the forecasting performance of models estimated on samples starting in 2004:1 could be able to assess the predictive power of the GI, but it would be of little practical relevance if models estimated on the longer sample were better at predicting unemployment rate dynamics.

Within the three groups we estimate pure time series $AR(p)$ and $ARMA(p, q)$ models, with at most 2 lags for p and q , for a total of four models ($AR(1)$, $AR(2)$, $ARMA(1,1)$ and $ARMA(2,2)$).

In addition, we augment these basic specifications with exogenous leading indicators, i.e. ARMAX(p, q):

$$\phi(L)u_t = \mu + x_t'\beta + \theta(L)\varepsilon_t \quad (1)$$

where x_t' is a vector with a first column of ones and one or more columns of leading indicators. These indicators should help improving the predictions of the US unemployment rate.

In particular, we use as exogenous variables (both on the short and the long sample) the monthly IC, i.e. IC_t , their weekly levels ($IC_{w1,t}$, $IC_{w2,t}$, $IC_{w3,t}$, and $IC_{w4,t}$) and their lags up to the second. We then estimated the same models for the short sample using the monthly average of the GI (G_t), its weekly values ($G_{w1,t}$, $G_{w2,t}$, $G_{w3,t}$, and $G_{w4,t}$) and their lags up to the second. Additionally, we augmented the four models with both leading indicators combined at the same frequency either monthly or weekly, at the same month t and for the previous months up to the second. Finally, the four models are estimated with both indicators, IC and the GI, both monthly and for each week. All these models are estimated adding seasonal multiplicative factors.⁷ In Table 4, we summarize all the groups of models within the short and the full sample.⁸

In our pseudo-out-of-sample exercise we consider the situation that real forecasters face when they produce their forecasts and the future values of the exogenous variables (x_t) need to be forecast. At any given date, we have run our forecasting horse-race using only the information that was really available at that time. Therefore, we have adopted simple ARMA models to predict x_t , so that we could use such predictions as inputs in our forecasting models. For robustness, we have considered different models⁹ but we present only those using an AR(1).

⁷In particular, we used a seasonal multiplicative autoregressive factor $SAR(12)$ for AR models and both an AR and MA seasonal $SMA(12)$ for ARMA models.

⁸In all our forecasting exercises we use a rolling window. However we have also performed our forecasting horse-race using a recursive scheme. The results are similar to those with a rolling scheme and are not reported for the sake of brevity, but they are available upon request.

⁹We have adopted an AR(1), AR(2), ARMA(1,1) and ARMA(2,2) and the results are quite similar.

4 Out-of-Sample Forecasting Comparison

When we perform an out-of-sample forecasting horse-race comparing numerous models it is extremely important to assess which model has the highest forecast accuracy with respect to a given benchmark or overall.

In Table 5 we present the mean squared errors (MSE), the Diebold and Mariano (DM) (1995) test of equal forecast accuracy and the Harvey et al. (HLN) (1998) test of forecast encompassing for the 15 best forecasting models of $u_t - u_{t-1}$, with forecast horizon from 1 to 3 months.¹⁰ For each forecast horizon the column labeled “Rank” gives the rank of each model in terms of lowest MSE. The first column labeled ‘n.’ denotes the number of the model. For the complete list of models see Table A.5 in the Appendix. We notice that for all forecast horizons the best model (i.e. the model with the lowest MSE out-of-sample) always includes the GI as the exogenous variable. In particular, the $ARX(1) - G_t$ (model #261), a standard AR(1) model with the average monthly GI, is the best model when forecasting both one and two months ahead. By the same token, the $ARMAX(1, 1) - G_t - SA$ (model #398), a standard ARMA(1,1) model with the average monthly GI plus a multiplicative seasonal factor, has the best performance among the three-month-ahead forecasts. It is important to notice that, at all forecast horizons, the best fifteen models always include the GI as an independent variable, in some cases in combination with the IC. Anyway, at one step ahead, the best 3 models include the GI only as an exogenous variable (thus not including IC). The same is also true for the two-step-ahead horizon (the best 5 models include only GI) and, even more, at the three-step-ahead horizon where the best 11 models include only our preferred leading indicator. Table 5 also reports the best models estimated over the full and the short sample without the GI. The reader can notice that for 1-month-ahead forecasts the best model without the GI over the full sample ranks 73rd, while the same model over the short sample ranks 197th. For 2- and 3-month-ahead forecasts these models without the GI rank higher than 173rd.

¹⁰Additional estimates for u_t and $\log(u_t)$ can be found in tables A.7 and A.8 of the Appendix.

The literature on US unemployment forecasting has thus far only considered the ratios of the mean squared errors between a competitor model and a benchmark model to evaluate each model forecast ability. Nevertheless, after the seminal papers by Diebold and Mariano (1995) and West (1996), the community of forecasters has increasingly understood the importance of correctly testing for out-of-sample equal forecast accuracy. West (2006) provides a recent survey of the tests of equal forecast accuracy, while Busetti et al. (2009) provide extensive Monte Carlo evidence on the best tests of equal forecast accuracy or forecast encompassing to be used in any specific framework (nested or non-nested models). To provide a more formal assessment of the forecasting properties of each model in our horse-race, we use the best model in terms of lowest MSE as the benchmark model and perform two tests. The first is a two-sided DM test for the null of equal forecast accuracy between the benchmark and the competitor and a two-sided HLN test, to assess whether the benchmark model forecast encompasses the competitor.¹¹ Recall that a benchmark model forecast encompasses the k -th competitor model if the former cannot be significantly improved upon by a convex forecast combination of the two. In other words, the benchmark forecast encompasses the competitor if this latter model does not provide any additional information for predicting. We use the two-sided version of these tests because some models are nested and others are non-nested making the direction of the alternative hypothesis unknown. Using the two-sided version of the tests we can thus compare both nested and non-nested models, as is our case where the exogenous variable often differs from one model to another and only a subset of models are really nested. Furthermore, we use both the DM and the HLN because they can be compared to standard critical values of the Gaussian distribution and Busetti et al. (2009) show

¹¹The DM test is based on the loss differential between the benchmark (model 0) and the k -th competitor, i.e. $d_t = e_{0,t}^2 - e_{k,t}^2$. To test the null of equal forecast accuracy $H_0 : E(d_t) = 0$, we employ the DM statistic $DM = P^{1/2}\bar{d}/\hat{\sigma}_{DM}$, where \bar{d} is the average loss differential, P is the out-of-sample size, and $\hat{\sigma}_{DM}$ is the square-root of the long-run variance of d_t . The HLN test analyzes the null $H_0 : E(f_t) = 0$, where $f_t = e_{0,t}(e_{0,t} - e_{k,t})$. The HLN test statistic is $HLN = P^{1/2}\bar{f}/\hat{\sigma}_{HLN}$, where \bar{f} is the average of the forecast error differential multiplied by the forecast error of the benchmark model, P is the out-of-sample size and $\hat{\sigma}_{HLN}$ is the square root of the long-run variance of f_t . Both tests are distributed as a Gaussian under the null.

that the HLN test is rather powerful both in a nested and non-nested framework when compared to other more complicated tests with non-standard distributions.

From Table A.5 in the Appendix we can see that the best model in terms of the lowest MSE always beats the competitors estimated on the full sample in predicting the unemployment rate in first differences. According to the standard DM test we can reject the null of equal forecast accuracy at 10% for 1- and 2-month-ahead forecast horizons. The same happens with the HLN test. At 10% we reject the null at the forecast horizons of 1 and 2 months. This means that our best model outperforms all those models that use the whole time series of unemployment and IC for the longest available time span, even though the former is estimated over a very short time window (38 months). When the benchmark is compared to models estimated on the short sample, both the DM and the HLN tests reject the null of equal forecast accuracy at 1-month ahead. However, they fail to reject the null for forecast horizons longer than 1-month.

In order to formally test the out-of-sample forecasting performance of the models using our suggested new leading indicator, we apply White’s (2000) “Reality Check” (RC) test. This test builds on Diebold and Mariano (1995) and West (1996) and involves examining whether the expected value of the forecast loss (e.g. the squared forecast error in the case of MSE) of one or several models is significantly greater than the forecast loss of a benchmark model. We adopt this test because in contrast to the previous ones, it tests for superior predictive ability rather than only for equal predictive ability. Furthermore, the RC test also allows us to account for the dependence among forecasting models that can arise when several models using the same data are compared in terms of predictive ability. Failing to do so can result in data-snooping problems, which occur when one searches a model extensively until a good match with the given data is found. White (2000) develops a test of superior unconditional predictive ability among multiple models accounting for this specification search. With this test we compare all the competitor models together against a benchmark. The null hypothesis is that all the models are no better than the benchmark, i.e., $H_0 : \max_{1 \leq k \leq L} E(f_k) \leq 0$, where $f_k = e_{0,t}^2 - e_{k,t}^2$ for MSE losses. This is

a multiple hypothesis, the intersection of the one-sided individual hypotheses $E(f_k) \leq 0$, $k = 1, \dots, L$. The alternative is that H_0 is false, that is, there exists a model which is superior to the benchmark. If the null hypothesis is rejected, there must be at least one model for which $E(f_k)$ is positive.¹² Hansen (2005) shows that White's Reality Check is conservative when a poor model is included in the set of L competing models. Hansen (2005) suggests using a studentized version of the RC test, suggesting the SPA test. We also tried the SPA test, but the two p-values are similar to the RC p-values and are not reported.

Table 6 reports the RC p-values for the best models against all the other models at each forecast horizon and for all the different transformations of the unemployment rate. In the Table we show the RC p-values for two different values of the probability parameter $q = (0.10, 0.50)$ and two different numbers of bootstrap replications $B = (2000, 5000)$. In boldface we report those RC p-values that are greater than the 5% significance level. We can notice that at this significance level we fail to reject the null hypothesis that none of the 519 competing models is better than our benchmark. Thus our best models with the GI have (almost always) superior predictive ability when compared to all the other models in our horse-race. However, we should acknowledge that these results must be interpreted with caution: we have a very short out-of-sample period and it is well known that the RC is undersized and has low power in small samples (see Hubrich and West, 2009).

¹²Suppose that $\sqrt{P}(\bar{f} - E(f)) \xrightarrow{d} N(0, \Omega)$ as $P(T) \rightarrow \infty$ when $T \rightarrow \infty$, for Ω positive semi-definite. White's (2000) RC test statistic for H_0 is formed as $\bar{V} = \max_{1 \leq k \leq L} \sqrt{P} \bar{f}_k$, where $\bar{f}_k = P^{-1/2} \sum_{t=R+1}^T \hat{f}_{k,t}$. However, as the null limiting distribution of \bar{V} is unknown, White (2000) showed that the distribution of $\sqrt{P}(\bar{f}^* - \bar{f})$ converges to that of $\sqrt{P}(\bar{f} - E(f))$, where \bar{f}^* is obtained from the stationary bootstrap of Politis and Romano (1994). By the continuous mapping theorem this result extends to the maximal element of the vector $\sqrt{P}(\bar{f}^* - \bar{f})$, so that the empirical distribution of $\bar{V}^* = \max_{1 \leq k \leq L} \sqrt{P}(\bar{f}_k^* - \bar{f}_k)$ may be used to compute the p-value of the test. This p-value is called the 'Reality Check p-value'.

5 Robustness checks

5.1 Nonlinear models

Most of the previous literature on unemployment forecasting in the US suggests using non-linear models to better approximate the long-term dynamic structure of its time series (see Montgomery et al., 1998 and Rothman, 1998). In particular, Montgomery et al. (1998) argue that Threshold Autoregressive (TAR) models can better approximate the unemployment rate dynamics especially during economic contractions, while linear ARMA models generally give a better representation of its short-term dynamics. To check the robustness of our best models which use the GI, we have also adopted some non-linear models that are typically used with the unemployment rate. We have estimated three non-linear time series models. The first is a self-exciting threshold autoregression (SETAR) model which takes the following form:

$$\begin{aligned} u_t = & [\phi_{01} + \phi_{11}u_{t-1} + \phi_{21}u_{t-2}] I(u_{t-1} \leq c) \\ & + [\phi_{02} + \phi_{12}u_{t-1} + \phi_{22}u_{t-2}] I(u_{t-1} > c) + \varepsilon_t \end{aligned} \quad (2)$$

where $I(\cdot)$ is the indicator function and c is the value of the threshold.

The SETAR models endogenously identify two different regimes given by the threshold variable u_{t-1} . In particular, following Rothman (1998) we adopted a SETAR model with two lags for each regime.

The second non-linear model used to forecast the unemployment rate is a logistic smooth transition autoregressive (LSTAR) model which is a generalization of the SETAR. The LSTAR model takes the form

$$\begin{aligned}
u_t = & [\phi_{01} + \phi_{11}u_{t-1} + \phi_{21}u_{t-2}] [1 - G(\gamma, c, u_{t-1})] \\
& + [\phi_{02} + \phi_{12}u_{t-1} + \phi_{22}u_{t-2}] G(\gamma, c, u_{t-1}) + \varepsilon_t
\end{aligned} \tag{3}$$

where $G(\gamma, c, u_{t-1}) = [1 + \exp(-\gamma \prod_{k=1}^K (u_t - c_k))]^{-1}$ is the logistic transition function, $\gamma > 0$ is the slope parameter set to zero for identification and c is the location parameter. In this model the change from one regime to the other is much smoother than in the SETAR model.

The third non-linear model employed to predict the US unemployment rate is an additive autoregressive model (AAR) of the following form

$$u_t = \mu + \sum_{i=1}^m s_i(u_{t-(i-1)d}) \tag{4}$$

where s_i are smooth functions represented by penalized cubic regression splines. The AAR model is a generalized additive model that combines additive models and generalized linear models. These models maximize the quality of prediction of a target variable from various distributions, by estimating a non-parametric function of the predictor variables which are connected to the dependent variable via a link function (see Hastie and Tibshirani, 1990). We have included this additional model to enlarge our out-of-sample comparison to non-parametric models which are found superior in predicting the US unemployment by Golan and Perloff (2004).

Panel C of Table 5 reports the MSE, the DM test and the HLN test for 1- to 3-month-ahead forecasts from these three non-linear models estimated only up to the second lag for the first differences of the US unemployment rate. At 1-month ahead the best non-linear model is the SETAR which ranks 258th, then the AAR (276) and the LSTAR (362). Thus, as previously found in the literature, non-linear models do not seem to be suitable for short-term forecasting. These non-linear models tend to fare better as soon as we forecast the unemployment rate at two and, in particular, at three months ahead, where their

rank ranges between the 24th and the 35th. We can thus conclude that our simple linear model using our preferred leading indicator (GI) also outperforms non-linear models, even though the gain tends to shrink as the forecast horizon increases.¹³

5.2 State level forecasts

As a further robustness check for the predictive properties of the GI, we estimated the same 520 models for each of the 51 states (including the District of Columbia), assessing the percentage of states for which the best model in terms of lower MSE is the one using the GI.

For the first-differenced series ($u_t - u_{t-1}$), the baseline in our forecast comparison, the percentage of the best models adopting the GI as a leading indicator ranges from 75% to 84% for the 1-step-ahead and the 3-step-ahead, respectively. When we use US unemployment rate in levels (u_t) as the dependent variable, the percentage of GI models with the lowest MSE out-of-sample ranges between 69% for the 2-step-ahead forecasts and 82% for the 3-step-ahead.

Finally, we test whether the aggregation of the 51 state models could improve the forecasting performance over the federal level benchmark. In particular, for each state we select the model with the lowest MSE and then aggregate the single state best forecasts using different weights. In Table 7 we compare the out-of-sample results of this aggregation with the benchmark model estimated at the federal (US) level, reported in the first row of each sub-panel as ‘best’ model. This model is characterized by the lowest MSE for the unemployment rate in first differences and in levels.

In particular, in the second row of Panel A of Table 7 we report the federal level forecasts obtained aggregating the state level estimates without weighting (simple average). In the third row, we weight the state level forecasts using the share of the labor force

¹³When we forecast the level u_t or the log-level $\log(u_t)$ of unemployment (see Tables A.7 and A.8 in the Appendix), these results hold only partially. In fact, non-linear models tend to rank poorly even at longer forecast horizon, thus showing that the linear models with Google clearly outperform nonlinear models even at longer horizons.

(employed plus unemployed) in state i on the total federal labor force. In the fourth row, this share is further weighted by the state i diffusion of the internet (See Table A.6 of the Appendix for descriptive statistics on internet diffusion among the entire population, among the active population aged 15-64, and among the 15-64 unemployed). The last row of each sub-panel is weighted by the share of unemployed combined with the 15-64 share of unemployed using internet. We define as internet diffusion in state i the share of individuals (active 15-64 individuals or unemployed 15-64 individuals according to the definition used) living in a household where at least an individual uses the internet.¹⁴

Forecasts obtained aggregating estimates of single state forecasts are inferior to the federal ones at all forecast horizons. Nevertheless, it is interesting to note that the gap between the best federal model and the aggregation of the 51 state models reduces as the forecast horizon increases, with MSEs being very close to the best federal-level forecasts in the three-step-ahead predictions. A more in-depth investigation of these patterns could be an interesting starting point for further research, but is beyond the scope of the present article.

5.3 Comparison with the Survey of Professional Forecasters

As an additional robustness check we compare the forecasts of our best model with the results of the Survey of Professional Forecasters (SPF), a quarterly survey of about 30 professionals, conducted by the Federal Reserve Bank of Philadelphia.¹⁵ The survey publishes estimates of the quarterly evolution of a set of macroeconomic variables approximately in the middle of the quarter.¹⁶

In Figure 5, we compare simple forecast errors for the median (SPF^{median}), the mean (SPF^{mean}) and the best individual forecast¹⁷ (SPF^{best}) of the SPF with those relative

¹⁴We calculate the weights using the results of the October 2007 supplement of the Current Population Survey (CPS). The exact question used for calculating the weights asks: *Do you (Does anyone) in this household use the Internet at any location? The possible answers are simply Yes/No.*

¹⁵<http://www.phil.frb.org/research-and-data/real-time-center/survey-of-professional-forecasters/>.

¹⁶The SPF is issued around the 15th of February, May, August and November.

¹⁷The best individual forecast is calculated ex-post once the real values for $u_t - u_{t-1}$ are known.

to the forecasts for each quarter obtained from a group of six best models. We define these best models as i) our best model overall (the one using the *GI*); ii) the best model among those not using the *GI* (*IC*) over the full sample; and iii) the best model among those not using the *GI* over the short sample (*IC_s*). To these three groups of best models we add three additional groups of non-linear models based on iv) the SETAR(2), v) the LSTAR(2) and vi) the AAR(2) model.

From each group we compute three series of quarterly forecasts. 1) $x^{1st-month}$ are the 1-month-ahead forecasts computed in the last month of each quarter before the one we want to forecast.¹⁸ The prediction for the whole quarter is equal to the forecast for the first month of the quarter. 2) $x^{2nd-month}$ are the 2-month-ahead forecasts computed in the last month of the quarter before, with the estimate for the whole quarter being equal to the estimate for the second, central, month. Both these forecasts are very conservative with respect to those of SPF, since the SPF is issued on the 15th of the second month of each reference quarter, thus around 45 days after our estimates are produced. Finally, 3) x^{Comb} are the quarterly forecasts computed as the average of the realized unemployment rate for the first month and the 1- and 2-month-ahead forecasts generated at the end of the first month of the reference quarter. These latter forecasts are less conservative because they use all the information available when the SPF is released. We thus expect that such forecasts should be at least as accurate as the SPF.

Does our model with Google outperform the professionals? It does, by a considerable margin, if we consider that it only uses a very short sample. In Table 8 we report the MSE for the nine best models and the three SPF forecasts over the period 2007Q2-2009Q2 along with the DM and the HLN tests where the benchmark is the model G^{Comb} , that is the model with the lowest MSE (in boldface). It is evident that the model including the *GI* outperforms all the three SPF forecasts, having a MSE lower by an order of magnitude. The DM test shows that the benchmark model is significantly better than all the other competitors using the first and second month forecasts, except for the less

¹⁸For example, if we want to forecast the quarterly unemployment rate for 2008Q2, at 2008.3 we compute the 1-month-ahead forecast from one of our three best models.

conservative forecasts x^{Comb} for which we reject the null hypothesis of equal forecast accuracy. Instead, the HLN test rejects the null that the benchmark model forecast encompasses the competitors, except for IC_s^{Comb} and $IC^{2nd-month}$. Figure 5 depicts the forecast errors from the best six models (those with the lowest MSE in Table 8) in addition to the mean and median SPF forecasts. It is rather clear that the model including the GI has the best performance in most periods, and in particular when the current recession worsened after the Lehman collapse in 2008Q4. We can see that the SPF and all the non-linear time series models tend to under-predict, whereas the linear models using either the IC or the GI tend to over-predict. While the models including the GI tend to give forecast errors that are close to zero, both the mean and median of the SPF tend to under-predict the real unemployment rate. This means that our simple linear ARMA models with the GI as a leading indicator outperform the predictions of the professional forecasters also during contractions, when the social impact of a high unemployment rate is even greater and the loss attached to high and positive forecast errors is maximal.¹⁹

6 Conclusions

Following the growing popularity of the internet as a job search tool and the increasing need of reliable and updated unemployment forecasts, especially during periods of economic downturn, in this paper we suggest the use of the Google index (GI), based on internet job-search performed through Google, as the best leading indicator to predict the US unemployment rate.

Popular time series specifications augmented with this indicator definitely improve their out-of-sample forecasting performance both at one-, two- and three-month horizons. Our results from the out-of-sample horse-race with more than five hundred linear and non-linear specifications show that the best models in terms of lowest MSE are always

¹⁹We have also performed the same robustness check for the forecasts using the level of the unemployment rate finding even more striking results that are unreported. In this case, all the model using GI outperform the SPF and, in particular, the best model is the $GI^{2-month}$.

those using the GI as the leading indicator. These models fare better also in comparison to other similar models estimated on a longer (or on the same) time span and using the initial claims (IC) as a leading indicator. These models outperform all the others both in terms of equal forecast accuracy and in terms of superior predictive ability. Our results are robust to various transformations of the dependent variable and are confirmed when assessing the predictive power of the GI in state-level forecasting. The best model including the GI also outperforms the forecasts released in the Survey of Professional Forecasters conducted by the Philadelphia Fed.

Notwithstanding its limited time availability (Google data start in January 2004) we believe that the GI should routinely be included in time series models to predict unemployment dynamics. It is easy to guess that the use of internet-based data will become widespread in economic research in the near future.

References

Askatas, Nikos and Zimmermann, Klaus F. 2009. “Google Econometrics and Unemployment Forecasting”, *IZA Discussion Paper*, (4201).

Buseti, Fabio, Marcucci, Juri and Veronese, Giovanni. 2009. “Comparing Forecast Accuracy: A Monte Carlo Investigation”, Bank of Italy, Discussion paper n. 723.

Choi, Hyunyoung and Varian, Hal. 2009. “Predicting Initial Claims for Unemployment Benefits”, *Google technical report*.

D’Amuri, Francesco. 2009. “Predicting unemployment in short samples with internet job search query data”, MPRA working paper n. 18403.

DeLong, James B. and Summers, Lawrence H. 1986. “Are Business Cycles Symmetrical?”, in *The American Business Cycle, Continuity and Changes*, ed. R. J. Gorton, Chicago: University of Chicago Press for NBER.

Diebold, Francis X. and Mariano, Roberto S. 1995. “Comparing Predictive Accuracy”, *Journal of Business & Economic Statistics*, 13: 253-263.

Elliott, Graham, Rothenberg, Thomas J., and Stock, James H. 1996. “Efficient Tests for an Autoregressive Unit Root,” *Econometrica*, 64: 813-836.

Ginsberg, Jeremy, Mohebbi, Matthew H., Patel, Rajan S., Brammer Lynnette, Smolinski, Mark S. and Brilliant Larry. 2009. “Detecting Influenza epidemics using Search Engine Query Data”, *Nature*, 457: 1012-1014.

Golan, Amos and Perloff, Jeffrey M. 2004. “Superior Forecasts of the U.S. Unemployment Rate Using a Nonparametric Method”, *The Review of Economics and Statistics*, February, 86(1): 433-438.

Hansen, Peter R. 2005. “A Test for Superior Predictive Ability”, *Journal of Business and Economic Statistics*, 23: 365-380.

Harvey, David I., Leybourne, Stephen J. and Newbold, Paul. 1998. “Tests for Forecast Encompassing”, *Journal of Business & Economic Statistics*, 16: 254-259.

Hastie, Trevor J., Tibshirani, Rob J. 1990. “Generalized Additive Models”, Chapman and Hall Ltd., London.

Hubrich, Kirstin, and West, Kenneth D. 2009. “Forecast Evaluation of Small

Nested Model Sets”, *ECB Working paper* n. 1030.

Koop, Gary, and Potter, Simon M. 1999. “Dynamic Asymmetries in U.S. Unemployment”, *Journal of Business and Economic Statistics*, 17(3): 298-312.

McQueen, Grant, and Thorley, Steven. 1993. “Asymmetric Business Cycle Turning Points”, *Journal of Monetary Economics*, 31: 341-362.

Montgomery, Alan L., Zarnowitz, Victor, Tsay, Ruey S. and Tiao, George C. 1998. “Forecasting the U.S. Unemployment Rate”, *Journal of the American Statistical Association*, June, 93(442): 478-493.

Neftci, Salih N. 1984. “Are Economic Time Series Asymmetric Over the Business Cycles?”, *Journal of Political Economy*, 85: 281-291.

Politis, Dimitris N. and Romano, Joseph P. 1994. “The Stationary Bootstrap”, *Journal of The American Statistical Association*, 89(428): 1303-1313.

Proietti, Tommaso. 2003. “Forecasting the US Unemployment Rate”, *Computational Statistics & Data Analysis*, 42: 451-476.

Rothman, Philip. 1998. “Forecasting Asymmetric Unemployment Rates”, *The Review of Economics and Statistics*, February, 80(1): 164-168.

Sichel, Daniel E.. 1993. “Business Cycle Asymmetry: A Deeper Look”, *Economic Enquiry*, 31: 224-236.

Suhoy, Tanya. 2009. “Query Indices and a 2008 Downturn ”, *Bank of Israel Discussion Paper (2009.06)*.

Wallis, Kenneth. 1987. “Time Series Analysis of Bounded Economic Variables”, *Journal of Time Series Analysis*, 8: 115-123.

West, Kenneth D. 1996. “Asymptotic inference about predictive ability”, *Econometrica*, 64: 1067-1084.

West, Kenneth D. 2006. “Forecast Evaluation”, 100-134, in *Handbook of Economic Forecasting*, Vol. 1, G. Elliott, C.W.J. Granger and A. Timmerman (eds), Amsterdam: Elsevier.

White, Halbert. 2000. “A Reality Check For Data Snooping”, *Econometrica*, 68(5): 1097-1126.

Table 1: Descriptive statistics: sample 2004:1-2009:6

	Mean	Median	Max	Min	Std. Dev.	Skew.	Kurt.	Jarque-Bera	Obs.
u_t	5.449	5.053	9.507	4.380	1.189	2.009	6.487	77.832***	66
$u_t - u_{t-1}$	0.058	0.026	0.539	-0.215	0.185	1.016	3.305	11.600***	66
$\log(u_t)$	1.676	1.620	2.252	1.477	0.187	1.610	5.029	39.819***	66
u_t^{logit}	-2.873	-2.933	-2.253	-3.083	0.200	1.637	5.121	41.838***	66
u_t^{LHP}	-0.019	-0.037	0.382	-0.191	0.139	1.087	3.905	15.239***	66
u_t^{LLD}	-0.140	-0.195	0.424	-0.340	0.184	1.550	4.900	36.372***	66
IC_t	1475.3	1337.5	2600.0	1152.0	365.3	2.035	5.983	70.037***	66
IC_{t-1}	1459.8	1337.5	2600.0	1152.0	343.7	2.209	6.948	96.539***	66
IC_{t-2}	1444.1	1337.5	2600.0	1152.0	317.2	2.382	8.093	133.767***	66
$IC_{w1,t}$	368.0	338.5	674.0	282.0	91.6	2.103	6.478	81.893***	66
$IC_{w1,t-1}$	363.9	338.5	674.0	282.0	85.8	2.287	7.588	115.427***	66
$IC_{w1,t-2}$	360.1	338.5	674.0	282.0	78.9	2.465	8.925	163.352***	66
$IC_{w2,t}$	367.4	333.5	660.0	288.0	90.2	2.061	6.243	75.629***	66
$IC_{w2,t-1}$	363.3	333.5	660.0	288.0	84.3	2.231	7.253	104.463***	66
$IC_{w2,t-2}$	359.7	333.5	660.0	288.0	78.7	2.433	8.601	151.386***	66
$IC_{w3,t}$	370.2	334.0	657.0	296.0	91.0	1.969	5.737	63.244***	66
$IC_{w3,t-1}$	366.6	334.0	657.0	296.0	86.2	2.134	6.633	86.396***	66
$IC_{w3,t-2}$	362.4	334.0	657.0	296.0	78.9	2.267	7.526	112.895***	66
$IC_{w4,t}$	369.7	330.5	645.0	284.0	95.8	1.891	5.340	54.400***	66
$IC_{w4,t-1}$	365.9	330.5	645.0	284.0	90.9	2.047	6.134	73.083***	66
$IC_{w4,t-2}$	361.9	330.5	645.0	284.0	84.4	2.193	7.021	97.361***	66
G_t	63.4	60.9	84.8	54.9	8.0	1.305	3.649	19.876***	66
G_{t-1}	63.2	60.6	84.8	54.9	7.8	1.388	3.968	23.402***	65
G_{t-2}	63.0	60.6	84.8	54.9	7.7	1.475	4.293	27.678***	64
$G_{w1,t}$	62.2	60.1	88.7	52.7	8.0	1.535	4.690	33.760***	66
$G_{w1,t-1}$	62.0	60.1	88.7	52.7	7.8	1.644	5.251	43.664***	66
$G_{w1,t-2}$	61.7	60.1	88.7	52.7	7.6	1.757	5.825	55.059***	65
$G_{w2,t}$	63.6	61.2	99.5	56.2	8.4	2.172	8.278	128.529***	66
$G_{w2,t-1}$	63.4	61.2	99.5	56.2	8.2	2.321	9.151	163.301***	66
$G_{w2,t-2}$	63.2	61.2	99.5	56.2	8.0	2.485	10.158	205.682***	65
$G_{w3,t}$	64.1	61.3	91.8	54.6	8.5	1.655	5.376	45.645***	66
$G_{w3,t-1}$	63.9	61.3	91.8	54.6	8.3	1.750	5.867	56.289***	66
$G_{w3,t-2}$	63.7	61.3	91.8	54.6	8.2	1.847	6.287	66.229***	65
$G_{w4,t}$	63.9	61.1	89.0	55.4	8.4	1.471	4.182	27.654***	66
$G_{w4,t-1}$	63.6	60.8	89.0	55.4	8.2	1.567	4.574	33.322***	65
$G_{w4,t-2}$	63.4	60.8	89.0	55.4	8.1	1.665	4.957	39.785***	64

Notes: u_t is the US unemployment rate in levels, $u_t - u_{t-1}$ are the first differences, $\log(u_t)$ is the unemployment rate in logs, $u_t^{logit} = \log(u_t/(1-u_t))$ is the logistic transformation suggested by Koop and Potter (1999), u_t^{LLD} is the log-linear de-trended unemployment rate and u_t^{LHP} is the HP-filtered series in log, both suggested by Rothman (1998). IC and G are the monthly initial claims and the monthly Google job search index used as leading indicators. The subscripts wj indicate the j^{th} week and $t-k$, $k = (0, 1, 2)$ is the time lag. ***, ** and * indicate rejection at 1, 5 and 10%, respectively.

Table 2: Correlations: sample 2004:1-2009:6

u_t	$d(u_t)$	$\log(u_t)$	u_t^{logit}	u_t^{LHP}	u_t^{LLD}	IC	IC_{-1}	IC_{-2}	IC_{w1}	IC_{w1}^{-1}	IC_{w2}^{-1}	IC_{w2}	IC_{w3}	IC_{w3}^{-1}	IC_{w3}	IC_{w3}^{-1}	IC_{w3}	
u_t	1																	
$u_t - u_{t-1}$	0.667	0.994	0.995	0.940	0.992	0.962	0.973	0.973	0.962	0.970	0.964	0.954	0.963	0.960	0.950	0.957	0.955	
$\log(u_t)$	0.674	1	0.674	0.542	0.662	0.711	0.673	0.673	0.700	0.664	0.604	0.699	0.681	0.596	0.706	0.640	0.632	
u_t^{logit}	0.994	0.674	1	1.000	0.999	0.951	0.956	0.951	0.948	0.951	0.941	0.941	0.947	0.938	0.939	0.940	0.933	
u_t^{LHP}	0.995	0.674	1	0.953	0.999	0.952	0.957	0.953	0.949	0.953	0.943	0.942	0.948	0.940	0.940	0.942	0.935	
u_t^{LLD}	0.940	0.542	0.953	1	0.962	0.844	0.864	0.873	0.842	0.859	0.860	0.832	0.854	0.856	0.842	0.858	0.866	
IC_t	0.992	0.662	0.999	0.962	1	0.942	0.949	0.946	0.940	0.945	0.936	0.933	0.940	0.932	0.931	0.934	0.929	
IC_{t-1}	0.962	0.711	0.951	0.952	0.844	1	0.977	0.975	0.990	0.965	0.940	0.992	0.965	0.938	0.992	0.961	0.938	
IC_{t-2}	0.973	0.673	0.956	0.957	0.864	0.949	0.977	0.975	0.982	0.988	0.961	0.972	0.991	0.960	0.961	0.991	0.957	
$IC_{w1,t}$	0.973	0.632	0.951	0.953	0.873	0.946	0.954	0.975	0.956	0.980	0.986	0.947	0.967	0.990	0.941	0.958	0.989	
$IC_{w1,t-1}$	0.962	0.700	0.948	0.949	0.842	0.940	0.990	0.982	1	0.969	0.946	0.989	0.971	0.941	0.968	0.968	0.937	
$IC_{w1,t-2}$	0.97	0.664	0.951	0.953	0.859	0.945	0.965	0.988	0.969	1	0.965	0.957	0.987	0.967	0.949	0.964	0.964	
$IC_{w2,t}$	0.964	0.604	0.941	0.943	0.860	0.936	0.940	0.961	0.946	0.965	1	0.936	0.950	0.985	0.922	0.944	0.957	
$IC_{w2,t-1}$	0.954	0.699	0.941	0.942	0.832	0.933	0.992	0.972	0.947	0.957	0.936	1	0.957	0.932	0.975	0.958	0.927	
$IC_{w2,t-2}$	0.963	0.681	0.947	0.948	0.854	0.940	0.965	0.991	0.971	0.987	0.950	0.957	1	0.950	0.950	0.972	0.952	
$IC_{w3,t}$	0.960	0.596	0.938	0.940	0.856	0.932	0.938	0.960	0.941	0.967	0.985	0.932	0.950	1	0.922	0.944	0.969	
$IC_{w3,t-1}$	0.950	0.706	0.939	0.940	0.842	0.931	0.992	0.961	0.968	0.949	0.922	0.975	0.950	0.922	1	0.942	0.930	
$IC_{w3,t-2}$	0.957	0.640	0.940	0.942	0.858	0.934	0.961	0.991	0.968	0.964	0.944	0.958	0.972	0.944	0.942	1	0.937	
$IC_{w4,t}$	0.955	0.632	0.933	0.935	0.866	0.929	0.938	0.957	0.937	0.964	0.957	0.927	0.952	0.969	0.930	0.937	1	
$IC_{w4,t-1}$	0.949	0.713	0.940	0.941	0.831	0.932	0.991	0.961	0.940	0.968	0.951	0.921	0.947	0.924	0.989	0.943	0.927	
$IC_{w4,t-2}$	0.962	0.679	0.947	0.948	0.852	0.940	0.980	0.990	0.981	0.965	0.948	0.975	0.967	0.942	0.965	0.987	0.939	
G_t	0.968	0.665	0.949	0.951	0.870	0.944	0.957	0.977	0.957	0.979	0.960	0.948	0.972	0.962	0.945	0.961	0.986	
G_{t-1}	0.851	0.745	0.866	0.865	0.706	0.854	0.902	0.862	0.885	0.847	0.818	0.890	0.848	0.809	0.886	0.840	0.794	
G_{t-2}	0.885	0.734	0.897	0.896	0.752	0.886	0.929	0.898	0.920	0.881	0.844	0.920	0.887	0.842	0.909	0.837	0.873	
$G_{w1,t}$	0.919	0.743	0.927	0.927	0.812	0.920	0.932	0.919	0.933	0.908	0.873	0.922	0.915	0.875	0.911	0.899	0.873	
$G_{w1,t-1}$	0.852	0.735	0.861	0.861	0.709	0.850	0.903	0.860	0.899	0.849	0.837	0.900	0.849	0.824	0.884	0.833	0.806	
$G_{w1,t-2}$	0.873	0.677	0.880	0.880	0.743	0.871	0.889	0.897	0.886	0.893	0.841	0.869	0.895	0.839	0.871	0.876	0.824	
$G_{w2,t}$	0.900	0.707	0.904	0.904	0.785	0.896	0.904	0.883	0.896	0.880	0.892	0.896	0.861	0.891	0.880	0.863	0.874	
$G_{w2,t-1}$	0.842	0.709	0.848	0.848	0.708	0.837	0.876	0.852	0.864	0.839	0.805	0.861	0.835	0.800	0.857	0.838	0.807	
$G_{w2,t-2}$	0.881	0.717	0.879	0.879	0.756	0.870	0.921	0.875	0.919	0.863	0.841	0.929	0.860	0.832	0.898	0.853	0.842	
$G_{w3,t}$	0.904	0.654	0.896	0.897	0.789	0.889	0.916	0.921	0.932	0.919	0.862	0.915	0.931	0.856	0.891	0.894	0.852	
$G_{w3,t-1}$	0.819	0.718	0.838	0.837	0.696	0.828	0.862	0.824	0.841	0.805	0.782	0.842	0.808	0.772	0.853	0.809	0.759	
$G_{w3,t-2}$	0.854	0.707	0.869	0.868	0.744	0.861	0.890	0.859	0.879	0.838	0.803	0.882	0.839	0.802	0.871	0.849	0.809	
$G_{w4,t}$	0.898	0.710	0.904	0.904	0.799	0.897	0.928	0.894	0.868	0.927	0.846	0.921	0.888	0.841	0.907	0.872	0.858	
$G_{w4,t-1}$	0.809	0.722	0.824	0.823	0.649	0.810	0.872	0.836	0.852	0.824	0.783	0.857	0.824	0.781	0.858	0.814	0.760	
$G_{w4,t-2}$	0.843	0.733	0.854	0.854	0.694	0.842	0.905	0.867	0.895	0.846	0.820	0.895	0.852	0.816	0.887	0.850	0.809	
	0.885	0.730	0.889	0.889	0.745	0.878	0.924	0.907	0.918	0.898	0.851	0.911	0.898	0.852	0.906	0.886	0.856	

Continued

Table 2: Correlations: sample 2004:1-2009:6 (Continued)

	IC^{w4}	IC_{-1}^{w4}	IC_{-2}^{w4}	G	G_{-1}	G_{-2}	G^{w1}	G_{-1}^{w1}	G_{-2}^{w1}	G^{w2}	G_{-1}^{w2}	G_{-2}^{w2}	G^{w3}	G_{-1}^{w3}	G_{-2}^{w3}	G^{w4}	G_{-1}^{w4}	G_{-2}^{w4}
u_t	0.949	0.962	0.968	0.851	0.885	0.919	0.852	0.873	0.900	0.842	0.881	0.904	0.819	0.854	0.898	0.809	0.843	0.885
$u_t - u_{t-1}$	0.713	0.679	0.665	0.745	0.734	0.743	0.735	0.677	0.707	0.709	0.717	0.654	0.718	0.707	0.710	0.722	0.733	0.730
$\log(u_t)$	0.940	0.947	0.949	0.866	0.897	0.927	0.861	0.880	0.904	0.848	0.879	0.896	0.838	0.869	0.904	0.824	0.854	0.889
u_t^{logit}	0.941	0.948	0.951	0.865	0.896	0.927	0.861	0.880	0.904	0.848	0.879	0.896	0.837	0.868	0.904	0.823	0.854	0.889
u_t^{LHP}	0.831	0.852	0.870	0.706	0.752	0.812	0.709	0.743	0.785	0.708	0.756	0.789	0.696	0.744	0.799	0.649	0.694	0.745
u_t^{LLD}	0.932	0.940	0.944	0.854	0.886	0.920	0.850	0.871	0.896	0.837	0.870	0.889	0.828	0.861	0.897	0.810	0.842	0.878
IC_t	0.991	0.980	0.957	0.902	0.929	0.932	0.903	0.889	0.904	0.876	0.921	0.916	0.862	0.890	0.872	0.905	0.924	0.924
IC_{t-1}	0.961	0.990	0.977	0.862	0.898	0.919	0.860	0.887	0.883	0.852	0.875	0.921	0.824	0.859	0.894	0.836	0.867	0.907
IC_{t-2}	0.940	0.958	0.989	0.823	0.859	0.892	0.835	0.852	0.896	0.824	0.855	0.875	0.787	0.824	0.868	0.791	0.832	0.872
$IC_{w,1,t}$	0.968	0.981	0.957	0.885	0.920	0.933	0.899	0.886	0.894	0.864	0.919	0.932	0.841	0.879	0.927	0.852	0.895	0.918
$IC_{w,1,t-1}$	0.951	0.965	0.979	0.847	0.881	0.908	0.849	0.893	0.880	0.839	0.863	0.919	0.805	0.838	0.882	0.824	0.846	0.898
$IC_{w,1,t-2}$	0.921	0.948	0.960	0.818	0.844	0.873	0.837	0.841	0.892	0.805	0.841	0.862	0.782	0.803	0.846	0.783	0.820	0.851
$IC_{w,2,t}$	0.971	0.975	0.948	0.890	0.920	0.922	0.900	0.869	0.896	0.861	0.929	0.915	0.842	0.882	0.921	0.857	0.895	0.911
$IC_{w,2,t-1}$	0.947	0.967	0.972	0.848	0.887	0.915	0.849	0.895	0.861	0.835	0.860	0.931	0.808	0.839	0.888	0.824	0.852	0.898
$IC_{w,2,t-2}$	0.924	0.942	0.962	0.809	0.842	0.875	0.824	0.839	0.891	0.800	0.832	0.856	0.772	0.802	0.841	0.781	0.816	0.852
$IC_{w,3,t}$	0.989	0.965	0.945	0.886	0.909	0.911	0.884	0.871	0.880	0.857	0.898	0.891	0.853	0.871	0.907	0.858	0.887	0.906
$IC_{w,3,t-1}$	0.943	0.987	0.961	0.840	0.880	0.899	0.833	0.876	0.863	0.838	0.853	0.894	0.809	0.849	0.872	0.814	0.850	0.886
$IC_{w,3,t-2}$	0.927	0.939	0.986	0.794	0.837	0.873	0.806	0.824	0.874	0.807	0.842	0.852	0.759	0.809	0.858	0.760	0.809	0.856
$IC_{w,4,t}$	1	0.963	0.944	0.913	0.933	0.930	0.897	0.899	0.902	0.890	0.908	0.894	0.879	0.924	0.888	0.909	0.909	0.930
$IC_{w,4,t-1}$	0.963	1	0.960	0.876	0.909	0.919	0.872	0.890	0.897	0.863	0.888	0.904	0.843	0.876	0.898	0.847	0.883	0.910
$IC_{w,4,t-2}$	0.944	0.960	1	0.832	0.874	0.903	0.835	0.866	0.886	0.842	0.865	0.887	0.799	0.842	0.883	0.802	0.842	0.888
G_t	0.913	0.876	0.832	1	0.982	0.936	0.957	0.935	0.913	0.930	0.896	0.837	0.978	0.955	0.908	0.984	0.978	0.959
G_{t-1}	0.933	0.909	0.874	0.982	1	0.967	0.951	0.954	0.931	0.944	0.927	0.890	0.954	0.977	0.953	0.953	0.983	0.978
G_{t-2}	0.930	0.919	0.903	0.936	0.967	1	0.919	0.938	0.942	0.923	0.929	0.912	0.912	0.937	0.968	0.900	0.935	0.967
$G_{w,1,t}$	0.897	0.872	0.835	0.957	0.951	0.919	1	0.888	0.890	0.833	0.898	0.851	0.924	0.907	0.933	0.964	0.964	0.912
$G_{w,1,t-1}$	0.899	0.890	0.866	0.935	0.954	0.938	1	0.888	0.880	0.818	0.824	0.892	0.912	0.921	0.891	0.923	0.929	0.963
$G_{w,1,t-2}$	0.907	0.892	0.886	0.913	0.931	0.942	0.899	0.880	1	0.913	0.914	0.814	0.905	0.908	0.919	0.869	0.918	0.926
$G_{w,2,t}$	0.890	0.863	0.842	0.930	0.944	0.923	0.833	0.918	0.913	1	0.893	0.823	0.918	0.965	0.909	0.896	0.915	0.962
$G_{w,2,t-1}$	0.908	0.888	0.865	0.896	0.927	0.929	0.898	0.824	0.914	0.893	1	0.888	0.854	0.914	0.963	0.845	0.891	0.911
$G_{w,2,t-2}$	0.894	0.904	0.887	0.837	0.890	0.912	0.851	0.892	0.814	0.823	0.888	1	0.774	0.847	0.911	0.798	0.836	0.885
$G_{w,3,t}$	0.879	0.843	0.799	0.978	0.954	0.912	0.924	0.912	0.905	0.918	0.854	0.774	1	0.935	0.877	0.955	0.965	0.935
$G_{w,3,t-1}$	0.897	0.876	0.842	0.955	0.977	0.937	0.894	0.921	0.908	0.965	0.914	0.847	0.935	1	0.931	0.920	0.952	0.964
$G_{w,3,t-2}$	0.924	0.898	0.883	0.908	0.953	0.908	0.907	0.891	0.919	0.909	0.963	0.911	0.877	0.931	1	0.859	0.916	0.950
$G_{w,4,t}$	0.888	0.847	0.802	0.984	0.953	0.900	0.933	0.923	0.869	0.896	0.845	0.798	0.955	0.920	0.859	1	0.957	0.939
$G_{w,4,t-1}$	0.909	0.883	0.842	0.978	0.983	0.935	0.964	0.929	0.918	0.915	0.891	0.836	0.965	0.952	0.916	0.957	1	0.955
$G_{w,4,t-2}$	0.930	0.910	0.888	0.959	0.978	0.967	0.912	0.963	0.926	0.962	0.911	0.885	0.935	0.964	0.950	0.939	0.955	1

Notes: u_t is the US unemployment rate in levels, $u_t - u_{t-1}$ are the first differences, $\log(u_t)$ is the unemployment rate in logs, $u_t^{logit} = \log(u_t/(1 - u_t))$ is the logistic transformation suggested by Koop and Potter (1999), u_t^{LLD} is the log-linear de-trended unemployment rate and u_t^{LHP} is the HP-filtered series in log, both suggested by Rothman (1998). IC and G are the monthly initial claims and the monthly Google job web search index used as leading indicators. Both the subscripts and superscripts wj indicate the j^{th} week and the subscripts $t - k$, $k = (0, 1, 2)$ is the time lag. ***, ** and * indicate rejection at 1, 5 and 10%, respectively.

Table 3: Unit Root tests for the US unemployment rate

Sample: 1967:1-2009:6			Sample: 2004:1-2009:6		
Variable	Test	Test stat.	Variable	Test	Test stat.
u_t	$DF - GLS^\mu$	-1.054	u_t	$DF - GLS^\mu$	-2.881***
	$DF - GLS^\tau$	-2.282		$DF - GLS^\tau$	-2.902*
$\log(u_t)$	$DF - GLS^\mu$	-0.901	$\log(u_t)$	$DF - GLS^\mu$	-2.792***
	$DF - GLS^\tau$	-2.190		$DF - GLS^\tau$	-2.797
u_t^{logit}	$DF - GLS^\mu$	-0.912	u_t^{logit}	$DF - GLS^\mu$	-2.801***
	$DF - GLS^\tau$	-2.203		$DF - GLS^\tau$	-2.804
u_t^{HPlog}	$DF - GLS^\mu$	-3.752***	u_t^{HPlog}	$DF - GLS^\mu$	-2.659***
	$DF - GLS^\tau$	-4.414***		$DF - GLS^\tau$	-2.523
u_t^{LLD}	$DF - GLS^\mu$	-1.344	u_t^{LLD}	$DF - GLS^\mu$	-2.823***
	$DF - GLS^\tau$	-2.190		$DF - GLS^\tau$	-2.797

Notes: The $DF - GLS^\mu$ test indicates the test where a constant is included as the exogenous regressor, while $DF - GLS^\tau$ is the test with a constant and trend included. The critical values at 1, 5, and 10% for the $DF - GLS^\mu$ test are -2.569 (-2.600), -1.941 (-1.946) and -1.616 (-1.614), respectively, for the full sample 1967.1-2009.6 (short sample 2004.1-2009.6). Instead, the critical values at 1, 5, and 10% for the $DF - GLS^\tau$ test are -3.48 (-3.709), -2.89 (-3.138) and -2.57 (-2.842), respectively, for the full sample 1967.1-2009.6 (short sample 2004.1-2009.6). ***, ** and * indicate rejection at 1, 5 and 10%, respectively.

Table 4: Forecasting Models: $\phi(L)y_t = \mu + x_t'\beta + \theta(L)\varepsilon_t$ for the unemployment rate

		Full Sample: 1967.1-2007.2				Short Sample: 2004.1-2007.2			
w/o LI	AR(1) #	AR(2) #	ARMA(1,1) #	ARMA(2,2) #	AR(1) #	AR(2) #	ARMA(1,1) #	ARMA(2,2) #	#
	u_{t-1}	u_{t-k}	$u_{t-1}, \varepsilon_{t-1}$	$u_{t-k}, \varepsilon_{t-k}$	u_{t-1}	u_{t-k}	$u_{t-1}, \varepsilon_{t-1}$	$u_{t-k}, \varepsilon_{t-k}$	1
w/ LI	x_t								
(t)									
IC	✓	1	✓	1	✓	1	✓	1	✓
IC _{wj}	✓	4	✓	4	✓	4	✓	4	✓
G	-	-	-	-	✓	1	✓	1	✓
G _{wj}	-	-	-	-	✓	4	✓	4	✓
IC, G	-	-	-	-	✓	1	✓	1	✓
IC _{wj} , G _{wj}	-	-	-	-	✓	5	✓	5	✓
(t-1)									
IC	✓	1	✓	1	✓	1	✓	1	✓
IC _{wj}	✓	4	✓	4	✓	4	✓	4	✓
G	-	-	-	-	✓	1	✓	1	✓
G _{wj}	-	-	-	-	✓	4	✓	4	✓
IC, G	-	-	-	-	✓	1	✓	1	✓
IC _{wj} , G _{wj}	-	-	-	-	✓	5	✓	5	✓
(t-2)									
IC	✓	1	✓	1	✓	1	✓	1	✓
IC _{wj}	✓	4	✓	4	✓	4	✓	4	✓
G	-	-	-	-	✓	1	✓	1	✓
G _{wj}	-	-	-	-	✓	4	✓	4	✓
IC, G	-	-	-	-	✓	1	✓	1	✓
IC _{wj} , G _{wj}	-	-	-	-	✓	5	✓	5	✓
j = 1, 4; k = 1, 2 - w/ or w/o SAR/SMA									
IC	✓	1	✓	1	✓	1	✓	1	✓
IC _{wj}	✓	4	✓	4	✓	4	✓	4	✓
G	-	-	-	-	✓	1	✓	1	✓
G _{wj}	-	-	-	-	✓	4	✓	4	✓
IC, G	-	-	-	-	✓	1	✓	1	✓
IC _{wj} , G _{wj}	-	-	-	-	✓	5	✓	5	✓

Notes: # indicates the number of models in each group. The subscript $w, j, j = 1, \dots, 4$ denotes the weekly leading indicators. A ✓ denotes that the model in that group adopts the row variable as a leading indicator.

Table 5: Forecasting US unemployment rate ($u_t - u_{t-1}$) in first differences. Best 15 models, best models without GI and non-linear models.

1-step ahead				2-step ahead				3-step ahead						
n.	Model	MSE Rank	DM	HLN	n.	Model	MSE Rank	DM	HLN	n.	Model	MSE Rank	DM	HLN
Panel A1: Best models														
261	ARX(1) - G_t	0.0166	1	-	261	ARX(1) - G_t	0.0157	1	-	398	ARMAX(1,1) - $G_t - SA$	0.0350	1	-
398	ARMAX(1,1) - $G_t - SA$	0.0167	2	0.060	464	ARMAX(2,2) - $G_t - SA$	0.0163	2	0.136	327	ARX(2) - G_t	0.0372	2	0.230
327	ARX(2) - G_t	0.0172	3	0.448	398	ARMAX(1,1) - $G_t - SA$	0.0166	3	0.177	332	ARX(2) - $G_t - SA$	0.0379	3	0.244
491	ARMAX(2,2) - $IC_{t-1} - G_{t-1}$	0.0177	4	0.328	327	ARX(2) - G_t	0.0172	4	0.633	261	ARX(1) - G_t	0.0382	4	0.308
305	ARX(1) - G_{t-2}	0.0179	5	0.616	266	ARX(1) - $G_t - SA$	0.0175	5	0.700	464	ARMAX(2,2) - $G_t - SA$	0.0382	5	0.295
464	ARMAX(2,2) - $G_t - SA$	0.0179	6	0.312	277	ARX(1) - $IC_t - G_t - SA$	0.0186	6	0.952	266	ARX(1) - $G_t - SA$	0.0383	6	0.299
371	ARX(2) - G_{t-2}	0.0181	7	0.614	332	ARX(2) - $G_t - SA$	0.0194	7	0.955	349	ARX(2) - G_{t-1}	0.0488	7	1.164
283	ARX(1) - G_{t-1}	0.0182	8	1.516	343	ARX(2) - $IC_t - G_t - SA$	0.0206	8	1.150	354	ARX(2) - G_{t-1}	0.0495	8	1.115
463	ARMAX(2,2) - $G_{w4,t} - SA$	0.0184	9	0.442	283	ARX(1) - G_{t-1}	0.0208	9	1.514	393	ARMAX(1,1) - G_t	0.0508	9	0.722
277	ARX(1) - $IC_t - G_t - SA$	0.0186	10	0.852	420	ARMAX(1,1) - $G_{t-1} - SA$	0.0217	10	0.981	288	ARX(1) - G_{t-1}	0.0510	10	1.142
271	ARX(1) - $IC_t - G_t$	0.0186	11	0.709	288	ARX(1) - $G_{t-1} - SA$	0.0220	11	1.402	283	ARX(1) - G_{t-1}	0.0513	11	1.217
266	ARX(1) - $G_t - SA$	0.0188	12	0.998	305	ARX(1) - G_{t-2}	0.0220	12	1.551	343	ARX(2) - $IC_t - G_t - SA$	0.0528	12	0.659
337	ARX(2) - $IC_t - G_t$	0.0191	13	0.799	349	ARX(2) - G_{t-1}	0.0222	13	1.915*	277	ARX(1) - $IC_t - G_t - SA$	0.0531	13	0.681
343	ARX(2) - $IC_t - G_t - SA$	0.0192	14	0.870	293	ARX(1) - $IC_{t-1} - G_{t-1}$	0.0233	14	1.989**	365	ARX(2) - $IC_{t-1} - G_{t-1} - SA$	0.0548	14	1.275
270	ARX(1) - $IC_{w4,t} - G_{w4,t}$	0.0192	15	0.778	299	ARX(1) - $IC_{t-1} - G_{t-1} - SA$	0.0234	15	1.392	265	ARX(1) - $G_{w4,t} - SA$	0.0555	15	0.938
Panel B1: Best models without Google														
122	ARMAX(2,2) - $IC_{w4,t-2}$	0.0234	73	2.491**	122	ARMAX(2,2) - $IC_{w4,t-2}$	0.0514	180	1.814*	122	ARMAX(2,2) - $IC_{w4,t-2}$	0.1406	191	1.309
133	ARMA(1,1)	0.0301	197	2.152**	234	ARMAX(2,2) - $IC_{w3,t} - SA$	0.0565	191	1.389	215	ARMAX(1,1) - $IC_{w4,t-1} - SA$	0.1294	173	1.748*
Panel C1: Non-linear models														
521	SETAR(2)	0.0332	258	2.434**	521	SETAR(2)	0.0388	97	1.053	521	SETAR(2)	0.0589	24	0.758
522	LSTAR(2)	0.0368	362	2.497**	522	LSTAR(2)	0.0447	140	1.190	522	LSTAR(2)	0.0620	30	0.790
523	AAR(2)	0.0342	276	2.337**	523	AAR(2)	0.0436	134	1.183	523	AAR(2)	0.0652	35	0.814
Panel B2: Best models without Google														
Panel C2: Non-linear models														
Panel B3: Best models without Google														
Panel C3: Non-linear models														

Notes: ***, ** and * indicate rejection at 1, 5 and 10%, respectively. This table reports the best 15 models in terms of MSE among the 523 estimated ones. The complete list of models and their forecasting performance is available in the Appendix (table A.5). SA indicates the model augmented with a multiplicative seasonal factor.

Table 6: Reality-Check p -values for testing the superior predictive ability of our best model (with Google Index) against all the other models

	u_t		$u_t - u_{t-1}$		$\log(u_t)$		u_t^{LLD}		u_t^{logit}		u_t^{HPlog}	
B	2000	5000	2000	5000	2000	5000	2000	5000	2000	5000	2000	5000
1-step	Benchmark=403	Benchmark=403	Benchmark=261	Benchmark=261	Benchmark=327	Benchmark=327	Benchmark=327	Benchmark=327	Benchmark=327	Benchmark=327	Benchmark=327	Benchmark=327
q=0.50	0.073	0.070	0.107	0.098	0.099	0.100	0.076	0.076	0.083	0.083	0.073	0.083
q=0.10	0.053	0.057	0.055	0.057	0.050	0.045	0.053	0.060	0.073	0.068	0.057	0.060
2-step	Benchmark=332	Benchmark=332	Benchmark=261	Benchmark=261	Benchmark=327	Benchmark=327	Benchmark=327	Benchmark=327	Benchmark=327	Benchmark=327	Benchmark=327	Benchmark=327
q=0.50	0.037	0.039	0.098	0.097	0.080	0.080	0.043	0.040	0.027	0.033	0.065	0.062
q=0.10	0.053	0.052	0.053	0.045	0.058	0.058	0.061	0.057	0.054	0.056	0.057	0.057
3-step	Benchmark=332	Benchmark=332	Benchmark=398	Benchmark=398	Benchmark=266	Benchmark=266	Benchmark=266	Benchmark=266	Benchmark=266	Benchmark=266	Benchmark=266	Benchmark=266
q=0.50	0.037	0.045	0.073	0.073	0.114	0.114	0.029	0.025	0.028	0.027	0.041	0.038
q=0.10	0.046	0.052	0.048	0.048	0.058	0.066	0.050	0.052	0.052	0.054	0.061	0.052

Notes: The null hypothesis of the Reality Check test is that none of the models beat the benchmark (i.e. our best model with Google index with the lowest MSE overall). B indicates the number of bootstrap replications and q is the probability parameter of the stationary bootstrap implemented to compute the Reality Check p -values. In boldface we indicate all the Reality Check p -values significant at 5% or more.

Table 7: Forecasts of the US unemployment rate aggregating single state level forecasts.

Panel A - Variable: $(u_t - u_{t-1})$	1-Step			2-Step			3-Step								
	MSE	Rk1	Rk2	DM	HLN	MSE	Rk1	Rk2	DM	HLN	MSE	Rk1	Rk2	DM	HLN
Best federal level model	0.0166	1	1	-	-	0.0157	1	1	-	-	0.0350	1	4	-	-
Aggregation of state level models															
Simple average	0.2845	7	525	5.30***	4.92***	0.3391	7	524	2.77***	2.31**	0.3966	7	510	1.99**	2.31**
Weighted avg (labor force)	0.0292	2	181	-0.13	2.68***	0.0310	2	48	-0.30	1.31	0.0411	2	7	-1.17	1.31
- labor force*(internet use among labor force)	0.0299	5	196	-0.06	2.75***	0.0314	3	51	-0.28	1.32	0.0413	3	8	-1.16	1.32
- labor force*(internet use among active)	0.0296	3	190	-0.09	2.69***	0.0318	4	56	-0.26	1.30	0.0423	4	9	-1.14	1.30
- labor force*(internet use among unemployed)	0.0298	4	194	-0.07	2.71***	0.0322	5	57	-0.25	1.31	0.0425	5	10	-1.13	1.31
- unemployed*(internet use among unemployed)	0.0917	6	519	2.33**	3.33***	0.0690	6	239	0.65	1.66*	0.0618	6	32	-0.53	1.66*
Panel B - Variable: u_t															
Model and weighting															
Best federal level model	0.0167	1	1	-	-	0.0169	1	7	-	-	0.0482	6	15	-	-
Aggregation of state level models															
Simple average	0.3000	7	526	5.29***	4.70***	0.3700	7	522	2.48**	2.15**	0.4560	7	514	1.83*	1.73*
Weighted avg (labor force)	0.0280	2	120	0.24	2.95***	0.0293	2	29	-1.23	0.37	0.0459	3	3	-1.06	0.54
- labor force*(internet use among labor force)	0.0283	3	131	0.26	2.98***	0.0294	3	30	-1.24	0.36	0.0454	2	2	-1.07	0.54
- labor force*(internet use among active)	0.0286	4	137	0.29	2.94***	0.0303	5	33	-1.21	0.38	0.0474	5	5	-1.04	0.55
- labor force*(internet use among unemployed)	0.0287	5	140	0.30	2.96***	0.0302	4	32	-1.21	0.38	0.0469	4	4	-1.05	0.56
- unemployed*(internet use among unemployed)	0.0709	6	513	2.06**	3.31***	0.0519	6	152	-0.65	1.41	0.0373	1	1	-1.16	0.70

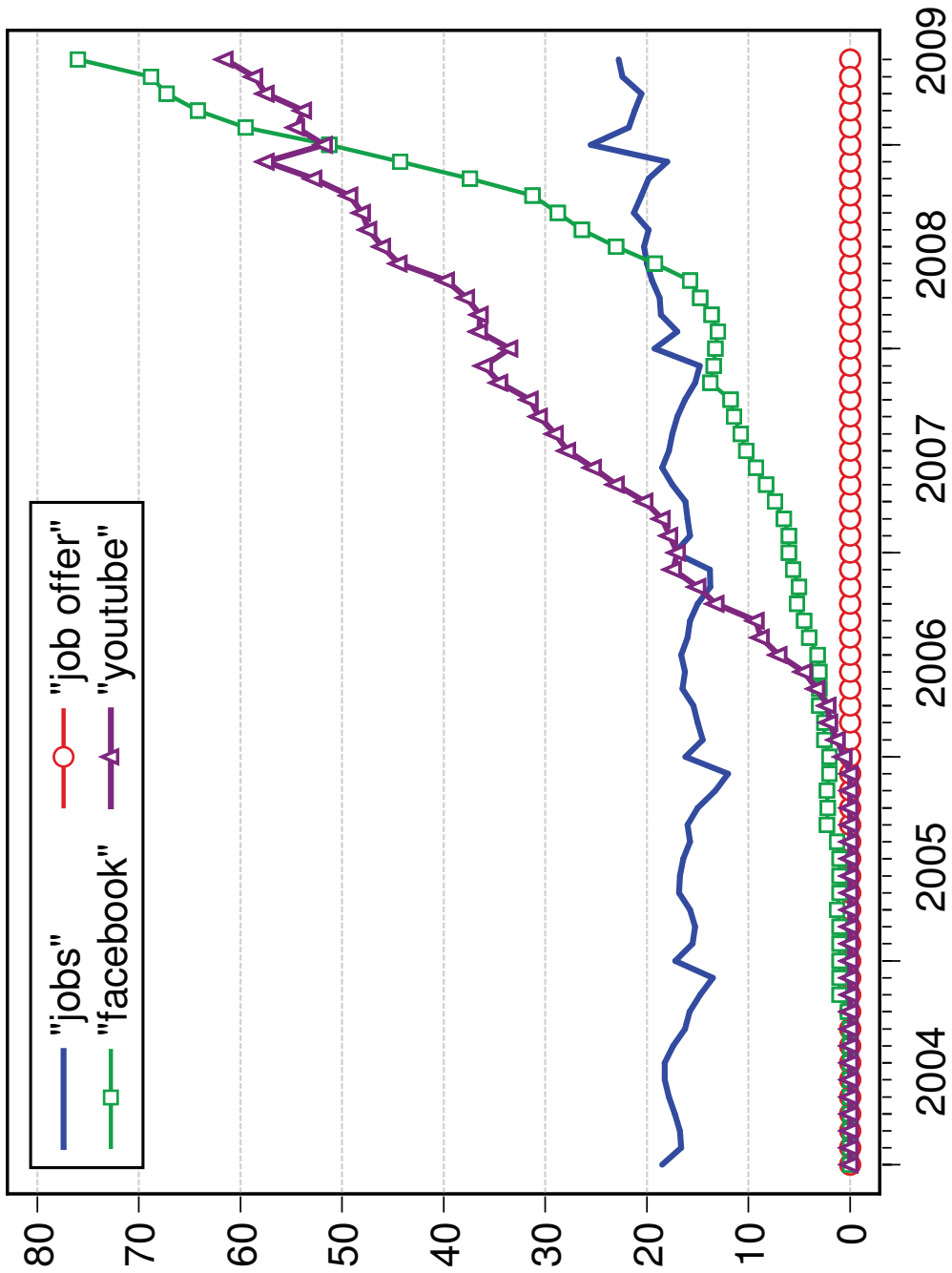
Notes: ***, ** and * indicate rejection at 1, 5 and 10%, respectively. The best federal level model is the model ranked first in the horse-race of table 5. Aggregation of state level models is made by taking the model with the lowest MSE for each state and then aggregating in a federal level forecast using a simple or weighed average as described in the table. Rk1 is the rank of each model within the table, while Rk2 is the rank of the model among all the models.

Table 8: Forecasts of the quarterly US unemployment: comparison of the best models with the Survey of Professional Forecasters.

	MSE	Rank	DM	HLN
<i>SPF^{best}</i>	1.373	21	1.911*	2.177**
<i>SPF^{mean}</i>	0.415	11	1.545	2.784***
<i>SPF^{med}</i>	0.360	7	1.317	2.892***
<i>G^{1st-month}</i>	0.530	15	-1.522	2.401**
<i>G^{2nd-month}</i>	0.419	12	1.724*	1.925*
<i>G^{Comb}</i>	0.082	1	-	-
<i>IC^{1st-month}</i>	0.893	17	-0.337	2.621***
<i>IC^{2nd-month}</i>	0.361	8	-0.919	1.457
<i>IC^{Comb}</i>	0.208	5	-2.012**	-1.875*
<i>IC_s^{1st-month}</i>	0.612	16	0.048	2.386**
<i>IC_s^{2nd-month}</i>	0.413	10	1.810*	1.759*
<i>IC_s^{Comb}</i>	0.218	6	1.306	1.239
<i>SETAR^{1st-month}</i>	1.123	19	2.881***	2.596***
<i>SETAR^{2nd-month}</i>	0.373	9	1.098	2.902***
<i>SETAR^{Comb}</i>	0.098	2	-1.401	2.587***
<i>LSTAR^{1st-month}</i>	1.228	20	2.558**	2.407**
<i>LSTAR^{2nd-month}</i>	0.433	14	1.550	2.723***
<i>LSTAR^{Comb}</i>	0.127	4	-1.265	2.315**
<i>AAR^{1st-month}</i>	1.060	18	2.630***	2.418**
<i>AAR^{2nd-month}</i>	0.432	13	1.768*	2.900***
<i>AAR^{Comb}</i>	0.102	3	-1.37	2.662***

Notes: In this table we compare the SPF one-quarter-ahead unemployment forecasts with similar forecasts generated from our best models for $u_t - u_{t-1}$, i.e. models n. 261, 261 and 398 for 1-, 2- and 3-month-ahead forecasts, respectively. The out-of-sample period is 2007.2-2009.6. *SPF^{best}* is the best individual forecaster in the survey, *SPF^{mean}* is the mean of the forecasts, while *SPF^{median}* is the median. Models *x^{1st-month}* are 1-month-ahead forecasts computed in the last month of the quarter before. Models *x^{2nd-month}* are 2-month-ahead forecasts computed in the last month of the quarter before. Both these forecasts are very conservative since the SPF is issued on the 15th of the second month of each reference quarter. Models *x^{Comb}* compute the quarterly forecast as the average of the realized unemployment rate for the first month and the 1- and 2-month-ahead forecasts generated at the end of the first month of the reference quarter. The model with Google is the best model overall, the model with *IC* is the best model without Google, while the models with subscript *IC_s* is the best model without Google in the short sample. SETAR, LSTAR and AAR are the corresponding non-linear models estimated over the full sample up to the second lag. In boldface we indicate the model with the minimum MSE, while in italics the next to the minimum MSE. The benchmark model for the DM and HLN tests is *G^{Comb}*. ***, ** and * indicate rejection at 1, 5 and 10%, respectively.

Figure 1: Relative incidence of keyword searches through Google



Notes: The figure depicts the relative incidence of the keyword searches 'jobs', adopted to construct our Google index along with the almost nil 'job offer', and the recently more popular 'facebook' and 'youtube' over the relevant sample 2004.1-2009.6.

Figure 2: Exact timing of monthly US Unemployment rate calculation

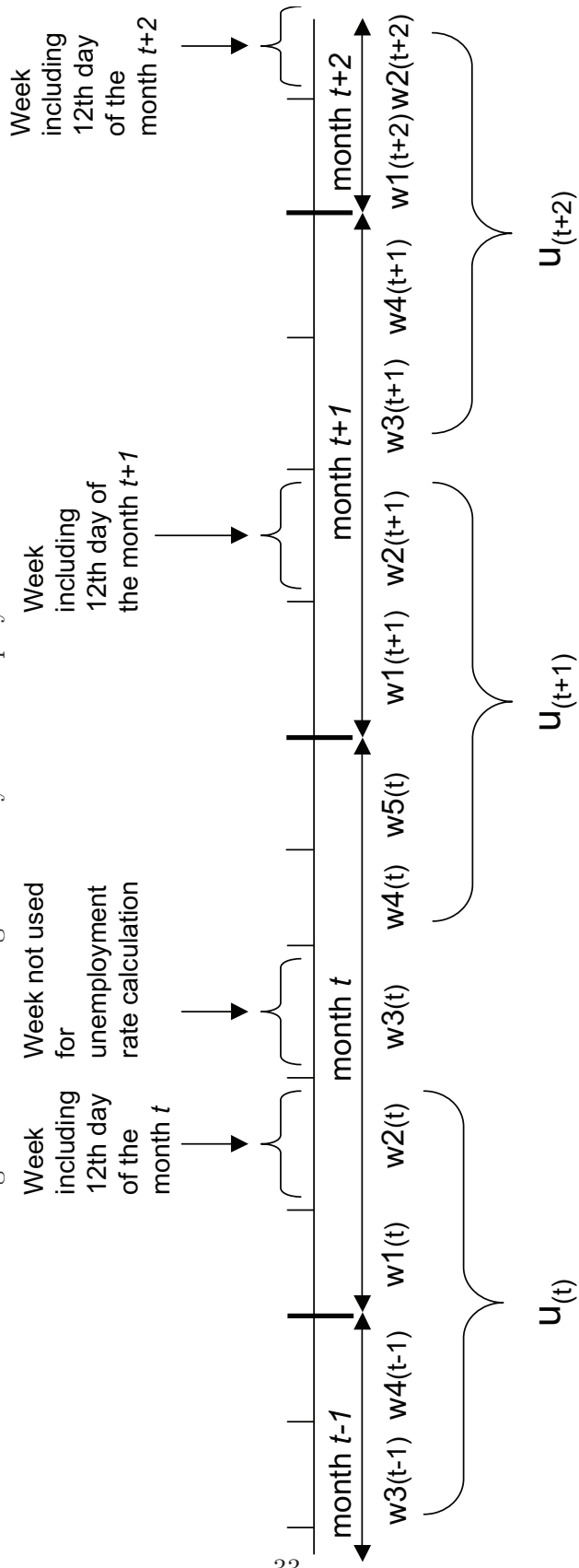
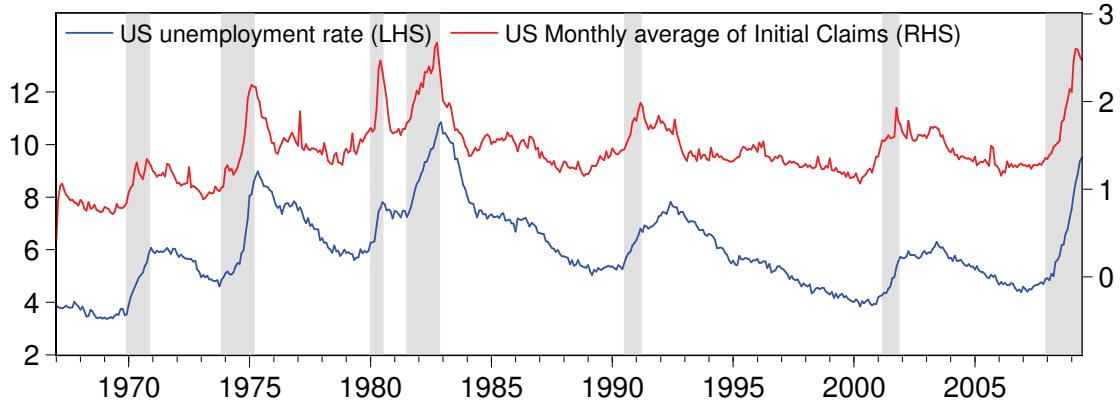
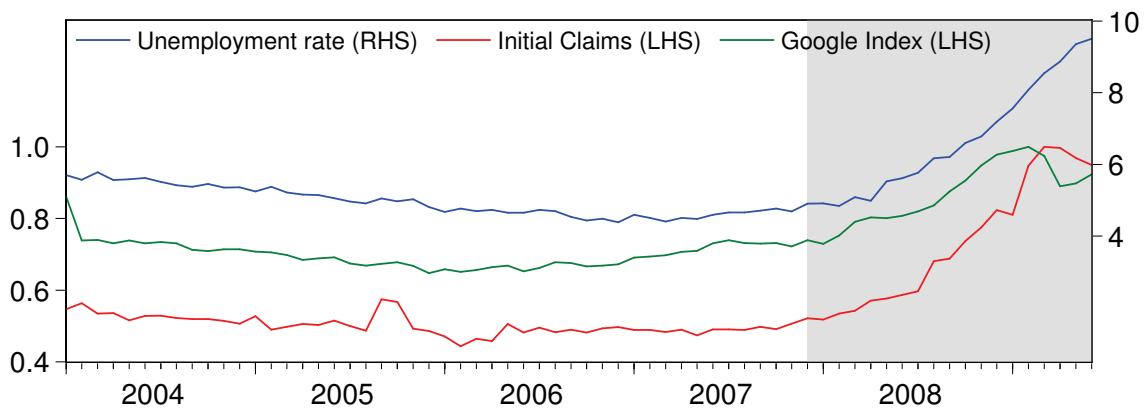


Figure 3: US Unemployment rate and Initial claims: Sample 1967:1-2009:6



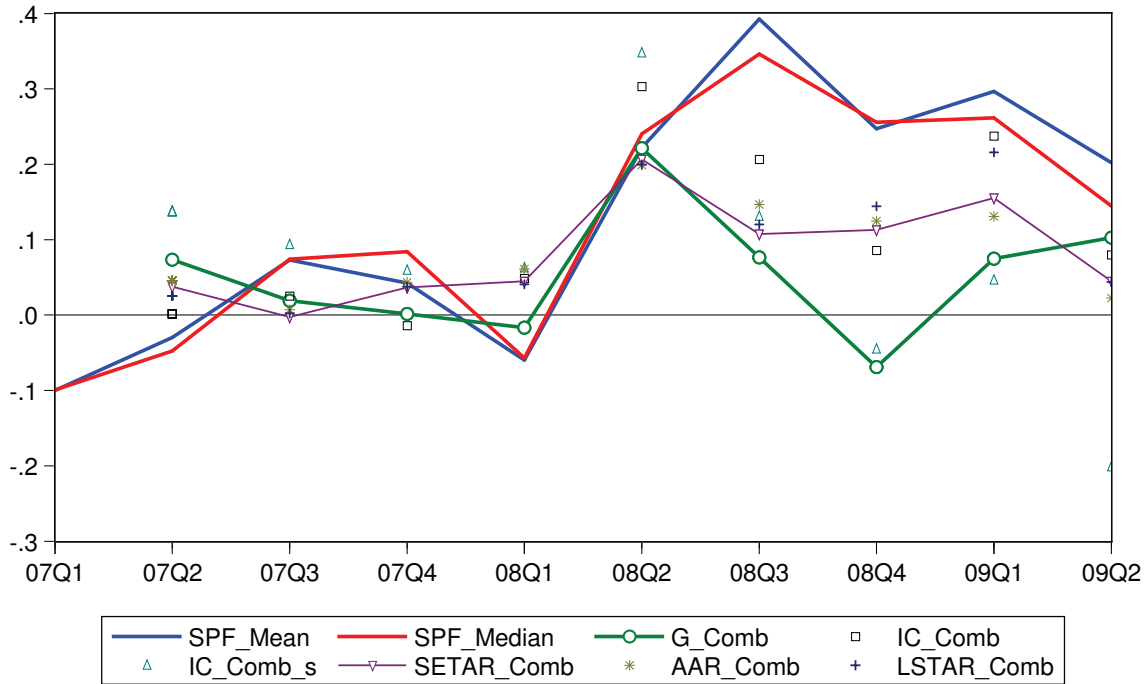
Notes: Shaded areas identify official NBER recessions.

Figure 4: US Unemployment rate, Initial claims and Google index: Sample 2004:1-2009:6



Notes: Shaded areas identify NBER recessions. The Initial claims are monthly averages rebased on their maximum over the sample 2004:1-2009:6. The Google index is the monthly average of Google 'job' searches rebased on their maximum value over the sample 2004:1-2009:6.

Figure 5: Forecast errors from quarterly forecasts of the US unemployment rate: comparison of the best models with the Survey of Professional Forecasters



Notes: In this table we compare the SPF one-quarter-ahead unemployment forecasts with similar forecasts generated from our best models for $u_t - u_{t-1}$, i.e. models n. 261, 261 and 398 for 1-, 2- and 3-month-ahead forecasts, respectively. The out-of-sample period is 2007.2-2009.6. SPF^{best} is the best individual forecaster in the survey, SPF^{mean} is the mean of the forecasts, while SPF^{median} is the median. Models $x^{1st-month}$ are 1-month-ahead forecasts computed in the last month of the quarter before. Models $x^{2nd-month}$ are 2-month-ahead forecasts computed in the last month of the quarter before. Both these forecasts are very conservative because the SPF is issued on the 15th of the second month of each reference quarter. Models x^{Comb} compute the quarterly forecast as the average of the realized unemployment rate for the first month and the 1- and 2-month-ahead forecasts generated at the end of the first month of the reference quarter. The model with Google (G) is the best model overall, the model with the Initial Claims (IC) is the best model without Google, while the models with subscript IC_s is the best model without Google in the short sample. SETAR, LSTAR and AAR are the corresponding non-linear models estimated over the full sample up to the second lag.

A Not-for-publication Appendix: Further Tables and Figures

Table A.1: Descriptive statistics of Initial Claims for the US and each single state - Full sample

	Mean	Median	Max	Min	Std. Dev.	Skew.	Kurt.	Jarque-Bera	Obs.
<i>ICUSA</i>	1430.7	1386.0	2673.0	415.0	344.2	0.720	4.511	92.580***	510
<i>ICAL</i>	25.9	25.7	44.5	15.7	5.0	0.705	4.077	35.530***	271
<i>ICAK</i>	7.0	7.1	11.4	4.4	1.0	0.067	4.064	12.984***	271
<i>ICAZ</i>	17.0	16.6	32.7	11.7	3.1	1.764	8.728	510.952***	271
<i>ICAR</i>	14.6	13.0	38.4	8.3	5.0	2.240	9.295	674.145***	271
<i>ICCA</i>	224.0	224.9	329.1	9.7	43.3	-0.096	4.218	17.177***	271
<i>ICCO</i>	11.4	11.1	23.9	7.0	2.7	1.544	7.658	352.690***	271
<i>ICCT</i>	18.4	17.5	33.6	10.7	4.0	0.775	3.419	29.122***	271
<i>ICDE</i>	4.1	4.0	9.5	1.7	1.2	0.664	3.721	25.792***	271
<i>ICDC</i>	2.2	2.3	6.7	1.0	0.9	0.881	5.139	86.733***	271
<i>ICFL</i>	40.8	37.5	121.8	24.0	14.9	2.679	12.069	1253.035***	271
<i>ICGA</i>	36.6	33.9	96.5	21.3	11.4	2.439	10.786	953.068***	271
<i>ICHI</i>	6.0	6.0	15.3	0.0	1.9	0.922	5.566	112.792***	271
<i>ICID</i>	8.7	8.3	17.1	5.9	1.6	2.329	10.999	967.554***	271
<i>ICIL</i>	57.9	55.2	112.8	40.0	11.1	2.037	9.489	662.804***	271
<i>ICIN</i>	27.3	26.9	74.9	14.6	9.6	2.078	9.228	633.094***	271
<i>ICIA</i>	13.0	12.2	42.5	7.5	4.6	3.544	19.880	3784.412***	271
<i>ICKS</i>	11.1	10.5	26.1	6.6	3.0	2.165	10.115	783.302***	271
<i>ICKY</i>	23.1	22.0	91.3	13.3	7.6	3.898	29.423	8569.968***	271
<i>ICLA</i>	17.2	14.9	215.0	8.6	15.9	9.793	111.022	136089.900***	271
<i>ICME</i>	7.5	6.7	21.7	4.5	2.5	1.590	6.735	271.631***	271
<i>ICMD</i>	18.4	17.5	34.1	12.7	3.7	1.439	5.547	166.823***	271
<i>ICMA</i>	32.5	30.4	55.1	22.1	7.1	0.925	3.118	38.788***	271
<i>ICMI</i>	71.9	69.1	160.6	42.0	20.1	1.512	6.333	228.666***	271
<i>ICMN</i>	19.9	18.8	41.9	12.1	4.6	1.607	7.050	301.850***	271
<i>ICMS</i>	14.4	14.0	60.0	9.1	4.4	5.483	52.606	29144.100***	271
<i>ICMO</i>	32.4	31.0	52.4	22.1	5.9	1.174	4.463	86.365***	271
<i>ICMT</i>	4.3	4.2	9.0	3.0	0.8	2.898	14.836	1961.225***	271
<i>ICNE</i>	61.6	56.8	120.5	27.9	17.3	0.921	3.541	41.615***	271
<i>ICNV</i>	2.4	2.3	7.5	1.3	0.6	4.297	31.933	10286.820***	271
<i>ICNH</i>	5.3	5.1	9.8	3.5	1.1	1.201	5.103	115.054***	271
<i>ICNJ</i>	4.0	3.8	8.5	1.8	1.3	0.982	3.877	52.242***	271
<i>ICNM</i>	42.8	42.5	65.2	30.6	6.3	0.722	4.058	36.178***	271
<i>ICNY</i>	4.9	4.8	9.8	0.1	1.0	1.070	13.165	1218.368***	271
<i>ICNC</i>	10.6	9.9	30.8	0.8	4.2	2.094	9.230	636.389***	271
<i>ICND</i>	86.2	84.2	139.7	54.2	14.2	1.082	4.665	84.186***	271
<i>ICOH</i>	53.3	50.7	110.4	31.4	13.6	1.576	6.525	252.464***	271
<i>ICOK</i>	9.7	9.3	20.9	5.2	2.5	1.161	4.908	101.958***	271
<i>ICOR</i>	28.1	26.6	57.5	17.6	6.4	1.716	7.104	323.298***	271
<i>ICPA</i>	89.0	86.6	164.0	61.7	14.2	2.339	12.413	1247.480***	271
<i>ICRI</i>	8.0	7.4	14.5	5.5	1.8	0.778	2.797	27.775***	271
<i>ICSC</i>	26.8	25.1	50.7	15.2	6.1	1.470	5.423	163.943***	271
<i>ICSD</i>	1.5	1.5	3.1	0.9	0.3	1.908	9.398	626.599***	271
<i>ICTN</i>	33.3	32.7	59.8	21.7	6.7	1.063	5.044	98.198***	271
<i>ICTX</i>	62.7	59.2	116.2	45.9	12.5	1.709	6.264	252.112***	271
<i>ICUT</i>	5.5	4.9	15.3	3.7	1.8	2.568	11.257	1067.685***	271
<i>ICVT</i>	3.3	3.2	6.7	-1.7	0.8	0.233	10.183	585.076***	271
<i>ICVA</i>	25.8	23.8	51.2	14.9	6.6	1.375	4.883	125.456***	271
<i>ICWA</i>	40.0	39.0	64.1	25.9	6.5	0.978	4.455	67.086***	271
<i>ICWV</i>	6.9	6.8	11.3	4.5	1.1	0.839	4.428	54.861***	271
<i>ICWI</i>	41.8	38.6	100.0	24.1	12.2	1.576	7.281	319.177***	271
<i>ICWY</i>	2.0	2.0	4.2	-0.9	0.5	0.399	7.310	216.985***	271

Notes: The subscript indicates the country (USA) or the state. For the US, the sample is 1967:1-2009:6, while for the single states the sample is 1986:12-2009:6. ***, ** and * indicate rejection at 1, 5 and 10%, respectively.

Table A.2: Descriptive statistics of Initial Claims for the US and each single state - Short sample

	Mean	Median	Max	Min	Std. Dev.	Skew.	Kurt.	Jarque-Bera	Obs.
<i>IC_{USA}</i>	1475.3	1337.5	2600.0	1152.0	365.3	2.035	5.983	70.037***	66
<i>IC_{AL}</i>	22.7	19.7	44.5	15.7	7.1	1.868	5.349	53.552***	66
<i>IC_{AK}</i>	6.7	6.5	8.8	5.8	0.7	1.256	4.085	20.594***	66
<i>IC_{AZ}</i>	17.1	15.5	32.7	12.1	4.8	1.841	5.720	57.626***	66
<i>IC_{AR}</i>	18.4	16.4	38.4	11.0	6.8	1.645	4.666	37.384***	66
<i>IC_{CA}</i>	194.1	180.7	310.7	144.8	44.0	1.516	4.170	29.042***	66
<i>IC_{CO}</i>	11.5	10.0	23.9	8.3	3.9	2.049	6.175	73.917***	66
<i>IC_{CT}</i>	18.1	16.8	28.6	15.2	3.5	1.959	5.642	61.433***	66
<i>IC_{DE}</i>	4.6	4.4	6.7	3.1	0.8	0.580	2.576	4.192	66
<i>IC_{DC}</i>	1.4	1.3	2.6	1.0	0.4	1.849	5.394	53.360***	66
<i>IC_{FL}</i>	53.0	43.7	121.8	32.7	22.4	1.527	4.342	30.588***	66
<i>IC_{GA}</i>	44.0	37.2	96.5	31.6	16.6	1.866	5.164	51.203***	66
<i>IC_{HI}</i>	5.5	4.8	10.6	3.4	1.9	1.478	3.973	26.631***	66
<i>IC_{ID}</i>	9.0	8.1	17.1	5.9	2.7	1.623	4.734	37.235***	66
<i>IC_{IL}</i>	61.1	56.1	112.8	49.1	15.3	2.334	7.508	115.798***	66
<i>IC_{IN}</i>	36.9	31.6	74.9	27.0	12.0	1.792	5.033	46.694***	66
<i>IC_{IA}</i>	16.3	13.6	42.5	10.7	7.3	2.328	7.620	118.347***	66
<i>IC_{KS}</i>	12.1	10.6	26.1	8.2	4.1	2.281	7.217	106.127***	66
<i>IC_{KY}</i>	26.4	22.9	54.1	16.0	9.1	1.785	5.460	51.681***	66
<i>IC_{LA}</i>	19.7	13.0	215.0	8.6	31.4	5.082	29.128	2161.375***	66
<i>IC_{ME}</i>	5.8	5.3	9.8	4.6	1.2	2.041	6.103	72.311***	66
<i>IC_{MD}</i>	18.7	16.8	34.1	13.5	4.9	1.859	5.383	53.641***	66
<i>IC_{MA}</i>	32.2	30.8	50.8	26.9	5.1	2.020	6.683	82.198***	66
<i>IC_{MI}</i>	78.0	71.1	160.6	59.4	20.8	2.260	7.784	119.106***	66
<i>IC_{MN}</i>	23.7	22.1	41.9	19.3	5.1	2.255	7.301	106.811***	66
<i>IC_{MS}</i>	13.4	11.2	60.0	9.1	7.5	4.708	27.244	1860.254***	66
<i>IC_{MO}</i>	32.4	30.5	52.4	24.4	6.8	1.732	5.348	48.170***	66
<i>IC_{MT}</i>	4.6	4.1	9.0	3.3	1.3	1.989	5.932	67.155***	66
<i>IC_{NE}</i>	58.5	52.6	120.5	41.9	17.7	2.080	6.451	80.330***	66
<i>IC_{NV}</i>	2.2	2.0	5.0	1.3	0.8	2.384	7.769	125.083***	66
<i>IC_{NH}</i>	6.0	5.8	9.8	4.6	1.2	1.491	4.912	34.516***	66
<i>IC_{NJ}</i>	4.3	3.9	8.5	3.4	1.3	2.098	6.204	76.644***	66
<i>IC_{NM}</i>	44.9	42.7	65.2	36.6	6.6	1.744	5.310	48.119***	66
<i>IC_{NY}</i>	4.9	4.6	9.8	3.1	1.5	2.017	6.443	77.331***	66
<i>IC_{NC}</i>	14.1	11.5	30.8	8.8	5.7	1.596	4.282	32.528***	66
<i>IC_{ND}</i>	86.7	81.1	139.7	66.7	16.3	1.836	5.641	56.268***	66
<i>IC_{OH}</i>	58.0	52.5	110.4	43.7	16.8	2.053	6.135	73.378***	66
<i>IC_{OK}</i>	9.6	8.5	20.9	6.0	3.2	1.843	6.115	64.061***	66
<i>IC_{OR}</i>	30.3	26.9	57.5	20.9	8.9	1.702	4.855	41.311***	66
<i>IC_{PA}</i>	96.4	88.9	164.0	80.4	20.7	2.218	7.035	98.878***	66
<i>IC_{RI}</i>	6.7	6.3	14.5	5.5	1.4	3.617	19.232	868.440***	66
<i>IC_{SC}</i>	27.2	24.3	50.3	19.9	7.3	1.867	5.329	53.242***	66
<i>IC_{SD}</i>	1.6	1.5	3.1	1.2	0.4	1.981	6.665	80.090***	66
<i>IC_{TN}</i>	29.3	26.7	59.8	21.7	9.1	2.096	6.423	80.532***	66
<i>IC_{TX}</i>	65.5	59.5	116.2	48.0	18.1	1.395	4.023	24.283***	66
<i>IC_{UT}</i>	6.4	5.5	15.3	4.0	2.8	1.877	5.522	56.261***	66
<i>IC_{VT}</i>	3.5	3.3	6.1	2.6	0.7	1.869	6.116	65.137***	66
<i>IC_{VA}</i>	24.2	21.6	46.9	17.1	7.2	1.971	5.823	64.636***	66
<i>IC_{WA}</i>	38.1	35.4	64.1	28.6	9.0	1.657	5.020	41.427***	66
<i>IC_{WV}</i>	6.1	5.7	11.2	4.6	1.4	2.398	8.386	143.050***	66
<i>IC_{WI}</i>	52.9	48.3	100.0	41.8	13.0	2.388	7.953	130.174***	66
<i>IC_{WY}</i>	1.8	1.6	4.2	1.0	0.7	2.196	7.323	104.444***	66

Notes: The subscript indicates the country (USA) or the state. The sample for the US and the single states is 2004:1-2009:6. ***, ** and * indicate rejection at 1, 5 and 10%, respectively.

Table A.3: Descriptive statistics of Google index for the US and each single state

	Mean	Median	Max	Min	Std. Dev.	Skew.	Kurt.	Jarque-Bera	Obs.
<i>GI_{USA}</i>	63.437	60.919	84.839	54.899	7.995	1.305	3.649	19.876***	66
<i>GI_{AL}</i>	61.650	59.597	80.778	50.670	7.585	1.148	3.599	15.490***	66
<i>GI_{AK}</i>	75.927	75.767	82.472	70.735	2.734	0.515	2.911	2.935	66
<i>GI_{AZ}</i>	52.178	49.486	70.849	43.314	7.400	1.116	3.208	13.821***	66
<i>GI_{AR}</i>	64.332	61.877	83.549	55.609	6.801	1.216	3.701	17.615***	66
<i>GI_{CA}</i>	38.437	38.214	48.748	33.787	2.968	0.986	4.177	14.498***	66
<i>GI_{CO}</i>	58.817	57.312	75.194	48.793	6.098	1.133	3.634	15.230***	66
<i>GI_{CT}</i>	48.856	46.474	61.727	43.010	5.317	1.186	3.086	15.481***	66
<i>GI_{DE}</i>	56.603	52.752	79.817	44.556	10.225	0.939	2.696	9.949***	66
<i>GI_{DC}</i>	52.161	51.126	67.535	45.360	4.867	1.431	4.793	31.379***	66
<i>GI_{FL}</i>	44.919	42.818	60.330	37.512	7.074	0.888	2.474	9.439***	66
<i>GI_{GA}</i>	50.258	47.891	66.686	42.319	6.325	1.276	3.497	18.583***	66
<i>GI_{HI}</i>	47.899	45.476	62.027	40.306	5.941	1.123	2.992	13.880***	66
<i>GI_{ID}</i>	59.562	56.952	81.643	49.267	8.350	1.136	3.312	14.462***	66
<i>GI_{IL}</i>	44.734	43.044	56.921	38.325	5.015	1.006	3.011	11.122***	66
<i>GI_{IN}</i>	48.955	47.443	63.651	41.945	5.166	1.408	4.224	25.921***	66
<i>GI_{IA}</i>	56.357	55.746	68.457	48.286	4.428	0.881	3.464	9.128**	66
<i>GI_{KS}</i>	55.156	53.236	70.825	48.312	5.565	1.335	3.936	22.006***	66
<i>GI_{KY}</i>	55.735	53.918	72.940	46.096	6.828	1.035	3.196	11.889***	66
<i>GI_{LA}</i>	53.601	53.125	70.330	42.478	6.356	0.850	3.393	8.374**	66
<i>GI_{ME}</i>	61.455	59.966	75.739	51.763	5.893	0.555	2.495	4.087	66
<i>GI_{MD}</i>	53.972	51.493	72.681	45.472	6.794	1.453	4.131	26.733***	66
<i>GI_{MA}</i>	39.725	38.155	50.671	35.021	4.375	1.187	3.234	15.636***	66
<i>GI_{MI}</i>	48.104	46.028	60.602	44.911	4.175	1.702	4.677	39.596***	66
<i>GI_{MN}</i>	48.357	46.906	63.063	42.128	4.852	1.422	4.302	26.898***	66
<i>GI_{MS}</i>	62.866	60.376	84.746	52.298	8.316	1.144	3.339	14.712***	66
<i>GI_{MO}</i>	48.143	46.225	61.602	42.127	5.217	1.407	3.831	23.683***	66
<i>GI_{MT}</i>	58.251	55.900	82.424	45.868	8.527	1.375	4.277	25.266***	66
<i>GI_{NE}</i>	55.852	54.692	70.379	48.175	4.879	1.279	4.109	21.360***	66
<i>GI_{NV}</i>	57.613	53.876	76.088	45.674	8.306	0.847	2.527	8.503**	66
<i>GI_{NH}</i>	58.316	55.653	80.347	48.795	7.540	1.145	3.479	15.041***	66
<i>GI_{NJ}</i>	45.386	43.264	60.192	39.252	5.654	1.372	3.618	21.745***	66
<i>GI_{NM}</i>	61.900	60.887	80.087	53.232	5.996	1.298	4.322	23.327***	66
<i>GI_{NY}</i>	39.346	38.168	48.891	34.967	3.992	1.086	3.113	12.999***	66
<i>GI_{NC}</i>	56.217	53.837	72.214	48.994	6.528	1.229	3.300	16.855***	66
<i>GI_{ND}</i>	60.669	61.089	69.816	50.779	4.085	0.049	2.712	0.255	66
<i>GI_{OH}</i>	49.950	47.640	64.258	42.536	5.391	1.331	3.733	20.964***	66
<i>GI_{OK}</i>	56.057	54.400	73.466	45.811	6.202	1.358	4.365	25.404***	66
<i>GI_{OR}</i>	48.891	48.318	58.723	42.633	4.501	0.653	2.543	5.264*	66
<i>GI_{PA}</i>	42.455	40.593	56.199	37.073	5.092	1.295	3.572	19.340***	66
<i>GI_{RI}</i>	53.536	49.963	69.884	45.062	7.272	0.907	2.413	9.995***	66
<i>GI_{SC}</i>	64.543	61.934	83.442	54.952	6.993	1.246	3.720	18.499***	66
<i>GI_{SD}</i>	62.359	60.382	84.677	50.115	8.074	1.147	3.817	16.295***	66
<i>GI_{TN}</i>	56.319	53.650	74.492	47.818	7.355	1.240	3.412	17.381***	66
<i>GI_{TX}</i>	47.254	46.202	63.223	39.614	6.267	1.140	3.415	14.771***	66
<i>GI_{UT}</i>	60.265	57.009	83.959	48.690	8.933	1.308	3.845	20.793***	66
<i>GI_{VT}</i>	57.103	56.193	72.158	48.735	5.157	0.982	3.713	12.009***	66
<i>GI_{VA}</i>	47.029	48.605	54.371	37.041	4.415	-0.912	2.739	9.330***	66
<i>GI_{WA}</i>	45.850	43.964	59.215	39.752	5.239	1.064	3.153	12.517***	66
<i>GI_{WV}</i>	59.866	58.874	77.394	47.424	5.855	0.840	3.928	10.127***	66
<i>GI_{WY}</i>	49.865	48.482	65.075	44.367	4.804	1.585	4.647	35.109***	66
<i>GI_{WY}</i>	60.580	58.442	80.550	51.443	6.646	1.468	4.479	29.731***	66

Notes: The subscript indicates the country (USA) or the state. For the US and all the states the sample is 2004:1-2009:6. ***, ** and * indicate rejection at 1, 5 and 10%, respectively.

Table A.4: Descriptive statistics of the unemployment rate for the US and each single state

	Mean	Median	Max	Min	Std. Dev.	Skew.	Kurt.	Jarque-Bera	Obs.
<i>urUSA</i>	5.621	5.523	10.849	2.548	1.512	0.599	3.467	50.792***	738
<i>urAL</i>	6.522	6.218	14.429	3.256	2.437	1.150	4.115	109.442***	402
<i>urAK</i>	8.041	7.611	11.494	5.894	1.485	0.405	1.939	29.866***	402
<i>urAZ</i>	5.979	5.774	11.480	3.592	1.635	1.278	4.597	152.176***	402
<i>urAR</i>	6.426	6.111	10.241	4.096	1.513	0.546	2.265	29.037***	402
<i>urCA</i>	7.034	6.833	11.611	4.726	1.594	0.621	2.667	27.710***	402
<i>urCO</i>	5.303	5.390	9.081	2.460	1.364	0.209	2.958	2.954	402
<i>urCT</i>	5.059	5.098	10.005	2.056	1.521	0.371	3.231	10.088***	402
<i>urDE</i>	5.021	4.351	8.428	2.891	1.737	0.616	1.881	46.440***	402
<i>urDC</i>	7.517	7.516	11.383	4.833	1.487	0.363	2.788	9.571***	402
<i>urFL</i>	6.039	5.869	10.553	3.325	1.582	0.459	2.673	15.893***	402
<i>urGA</i>	5.527	5.295	10.135	3.379	1.224	0.805	3.553	48.490***	402
<i>urHI</i>	4.671	4.748	10.170	2.192	1.589	0.657	3.642	35.805***	402
<i>urID</i>	5.821	5.494	9.412	2.778	1.453	0.384	2.972	9.894***	402
<i>urIL</i>	6.731	6.370	12.864	4.100	1.837	1.128	4.234	110.741***	402
<i>urIN</i>	5.879	5.343	12.849	2.577	2.248	1.139	3.883	99.980***	402
<i>urIA</i>	4.695	4.300	8.538	2.552	1.541	1.094	3.239	81.087***	402
<i>urKS</i>	4.587	4.473	7.404	2.938	0.812	0.608	3.968	40.470***	402
<i>urKY</i>	6.682	5.950	12.111	4.041	1.821	1.007	3.228	68.755***	402
<i>urLA</i>	7.220	6.591	12.856	3.176	2.391	0.797	2.682	44.269***	402
<i>urME</i>	5.713	5.370	9.001	2.987	1.516	0.401	2.170	22.303***	402
<i>urMD</i>	5.114	4.757	8.333	3.330	1.242	0.714	2.680	35.915***	402
<i>urMA</i>	5.511	5.276	10.941	2.655	1.811	0.652	2.729	29.707***	402
<i>urMI</i>	7.999	7.348	16.905	3.227	2.917	0.858	3.563	54.667***	402
<i>urMN</i>	4.855	4.711	9.021	2.475	1.300	0.883	4.125	73.453***	402
<i>urMS</i>	7.742	7.077	13.707	4.871	2.035	0.972	3.102	63.430***	402
<i>urMO</i>	5.740	5.600	10.476	2.593	1.468	0.868	4.256	76.932***	402
<i>urMT</i>	5.754	5.660	8.685	3.216	1.330	0.183	2.559	5.508*	402
<i>urNE</i>	3.469	3.143	6.849	2.159	0.951	1.055	3.638	81.454***	402
<i>urNV</i>	6.049	5.596	11.954	4.209	1.639	1.138	3.777	96.868***	402
<i>urNH</i>	4.332	3.953	7.743	1.870	1.480	0.704	2.540	36.762***	402
<i>urNJ</i>	6.077	5.874	10.644	3.502	1.736	0.640	2.737	28.593***	402
<i>urNM</i>	6.779	6.767	9.927	3.481	1.522	-0.100	2.491	5.004*	402
<i>urNY</i>	6.514	6.388	10.490	4.047	1.520	0.454	2.530	17.525***	402
<i>urNC</i>	5.449	5.269	11.101	3.099	1.590	1.187	4.662	140.655***	402
<i>urND</i>	4.098	4.012	6.867	2.511	0.965	0.523	2.392	24.482***	402
<i>urOH</i>	6.678	6.057	13.816	3.880	2.124	1.344	4.606	164.168***	402
<i>urOK</i>	5.239	5.030	9.400	2.714	1.503	0.605	2.745	25.641***	402
<i>urOR</i>	7.041	6.490	12.207	4.684	1.841	0.988	3.130	65.701***	402
<i>urPA</i>	6.444	5.857	12.902	4.039	1.869	1.217	4.535	138.732***	402
<i>urRI</i>	6.041	5.403	12.404	2.937	1.798	0.558	2.754	21.902***	402
<i>urSC</i>	6.161	6.024	12.078	3.083	1.664	1.130	5.033	154.748***	402
<i>urSD</i>	3.732	3.549	5.895	2.432	0.748	0.828	2.933	45.958***	402
<i>urTN</i>	6.382	5.789	12.361	3.791	1.899	1.478	4.740	197.094***	402
<i>urTX</i>	6.076	5.999	9.307	4.313	1.215	0.560	2.730	22.261***	402
<i>urUT</i>	4.891	4.698	9.735	2.423	1.471	0.887	3.765	62.510***	402
<i>urVT</i>	4.796	4.455	8.991	2.224	1.456	0.737	2.832	36.898***	402
<i>urVA</i>	4.539	4.473	7.846	2.188	1.215	0.273	2.754	6.022**	402
<i>urWA</i>	6.907	6.647	12.192	4.392	1.790	0.910	3.499	59.631***	402
<i>urWV</i>	8.424	7.695	18.197	4.090	3.288	0.927	3.311	59.150***	402
<i>urWI</i>	5.334	4.819	11.774	2.863	1.782	1.325	4.507	155.710***	402
<i>urWY</i>	4.944	4.693	10.090	1.898	1.631	0.930	3.647	64.999***	402

Notes: The subscript indicates the country (USA) or the state. For the US the sample is 1948:1-2009:6, while for the single states the sample is 1976:1-2009:6. ***, ** and * indicate rejection at 1, 5 and 10%, respectively.

Table A.5: Forecasting US unemployment rate ($u_t - u_{t-1}$) in first differences.

Model	MSE			DM			DLN		
	1-Step Rank	2-Step Rank	3-Step Rank	1-Step Rank	2-Step Rank	3-Step Rank	1-Step Rank	2-Step Rank	3-Step Rank
Model									
1 AR(1)	0.0564	507	0.1842	521	0.4270	516	3.329***	2.108**	1.810**
2 AR(1) - SA	0.0577	508	0.1894	522	0.4391	519	3.310***	2.119**	1.824*
3 AR(2)	0.0388	404	0.1063	454	0.2826	459	2.993***	1.959*	1.737*
4 AR(2) - SA	0.0395	421	0.1094	461	0.2905	466	3.043***	1.998**	1.755*
5 ARMA(1,1)	0.0354	310	0.0834	326	0.2048	320	2.530***	1.800*	1.625
6 ARMA(1,1) - SA	0.0357	329	0.0954	406	0.2339	402	2.577***	1.985**	1.783*
7 ARMA(2,2)	0.0314	229	0.0718	252	0.1833	258	2.314**	1.684*	1.583
8 ARMA(2,2) - SA	0.0324	252	0.0886	370	0.2172	367	2.564***	1.852**	1.760*
9 ARX(1) - IC _{w1,t}	0.0458	471	0.1365	480	0.3286	480	2.895***	2.072**	1.869*
10 ARX(1) - IC _{w2,t}	0.0454	465	0.1357	488	0.3256	478	2.913***	2.040**	1.942*
11 ARX(1) - IC _{w3,t}	0.0452	461	0.1303	483	0.3145	474	2.933***	2.174**	1.868*
12 ARX(1) - IC _{w4,t}	0.0418	441	0.1170	477	0.2843	461	2.805***	1.999**	1.756*
13 ARX(1) - IC _t	0.0439	449	0.1263	482	0.3044	471	2.857***	1.926*	1.823**
14 ARX(1) - IC _{w1,t} - SA	0.0470	476	0.1418	494	0.3423	485	2.961***	2.094**	1.882*
15 ARX(1) - IC _{w2,t} - SA	0.0465	474	0.1407	493	0.3387	483	2.971***	2.063**	1.881*
16 ARX(1) - IC _{w3,t} - SA	0.0462	472	0.1348	487	0.3261	479	2.979***	2.183**	1.961**
17 ARX(1) - IC _{w4,t} - SA	0.0424	444	0.1204	480	0.2937	468	2.836***	2.207**	2.002**
18 ARX(1) - IC _t - SA	0.0448	459	0.1307	484	0.3160	475	2.902***	2.118**	1.929*
19 ARX(1) - IC _{w1,t-1}	0.0487	485	0.1493	502	0.3568	493	3.038***	2.087**	1.847*
20 ARX(1) - IC _{w2,t-1}	0.0481	481	0.1471	501	0.3510	490	3.037***	2.067**	1.850*
21 ARX(1) - IC _{w3,t-1}	0.0484	483	0.1456	499	0.3476	489	3.066***	2.152**	1.890*
22 ARX(1) - IC _{w4,t-1}	0.0453	463	0.1328	485	0.3193	476	2.971***	2.171**	1.934*
23 ARX(1) - IC _{t-1}	0.0474	479	0.1422	496	0.3397	484	3.019***	2.113**	1.882*
24 ARX(1) - IC _{w1,t-1} - SA	0.0504	496	0.1565	510	0.3740	501	3.111***	2.118**	1.861*
25 ARX(1) - IC _{w2,t-1} - SA	0.0498	490	0.1543	507	0.3683	499	3.112***	2.100**	1.867*
26 ARX(1) - IC _{w3,t-1} - SA	0.0501	493	0.1529	506	0.3649	498	3.131***	2.171**	1.907*
27 ARX(1) - IC _{w4,t-1} - SA	0.0469	475	0.1398	492	0.3364	482	3.044***	2.186**	1.937*
28 ARX(1) - IC _{t-1} - SA	0.0491	487	0.1494	503	0.3571	494	3.091***	2.136**	1.890*
29 ARX(1) - IC _{w1,t-2}	0.0462	473	0.1359	509	0.3768	506	2.653***	1.954*	1.840*
30 ARX(1) - IC _{w2,t-2}	0.0446	455	0.1311	504	0.3648	497	2.671***	1.875*	1.770*
31 ARX(1) - IC _{w3,t-2}	0.0501	494	0.1517	505	0.3622	496	3.123***	2.094**	1.867*
32 ARX(1) - IC _{w4,t-2}	0.0446	456	0.1376	490	0.3342	481	3.066***	2.159**	1.917*
33 ARX(1) - IC _{t-2}	0.0440	450	0.1421	495	0.3449	487	2.707***	1.922**	1.792*
34 ARX(1) - IC _{w1,t-2} - SA	0.0473	478	0.1605	513	0.3877	509	2.714***	1.977**	1.773*
35 ARX(1) - IC _{w2,t-2} - SA	0.0455	467	0.1558	508	0.3757	502	2.734***	1.899*	1.789*
36 ARX(1) - IC _{w3,t-2} - SA	0.0517	499	0.1592	512	0.3790	508	3.164***	2.114**	1.880*
37 ARX(1) - IC _{w4,t-2} - SA	0.0457	470	0.1432	497	0.3475	488	3.112***	2.171**	1.921*
38 ARX(1) - IC _{t-2} - SA	0.0448	460	0.1466	500	0.3552	492	2.745***	1.939*	1.809*
39 ARX(2) - IC _{w1,t}	0.0357	328	0.0940	402	0.2516	426	2.675***	1.913*	1.767*
40 ARX(2) - IC _{w2,t}	0.0354	309	0.0931	397	0.2488	420	2.679***	1.893*	1.766*
41 ARX(2) - IC _{w3,t}	0.0354	312	0.0901	383	0.2420	413	2.689***	1.975**	1.832*
42 ARX(2) - IC _{w4,t}	0.0333	268	0.0826	318	0.2222	381	2.532***	1.988**	1.856*
43 ARX(2) - IC _t	0.0347	289	0.0885	369	0.2368	407	2.621***	1.931*	1.806*
44 ARX(2) - IC _{w1,t} - SA	0.0364	359	0.0970	417	0.2603	437	2.763***	1.948*	1.786*
45 ARX(2) - IC _{w2,t} - SA	0.0361	345	0.0961	410	0.2574	433	2.764***	1.927*	1.785*
46 ARX(2) - IC _{w3,t} - SA	0.0361	343	0.0931	396	0.2503	424	2.766***	2.001**	1.844*
47 ARX(2) - IC _{w4,t} - SA	0.0339	274	0.0853	341	0.2300	397	2.603***	2.009**	1.869*
48 ARX(2) - IC _t - SA	0.0353	307	0.0914	390	0.2452	417	2.700***	1.957*	1.818*
49 ARX(2) - IC _{w1,t-1}	0.0371	372	0.0995	428	0.2665	443	2.839***	1.930*	1.749*
50 ARX(2) - IC _{w2,t-1}	0.0368	368	0.0982	423	0.2629	440	2.823***	1.918*	1.743*
51 ARX(2) - IC _{w3,t-1}	0.0370	369	0.0976	418	0.2617	438	2.846***	1.956*	1.788*
52 ARX(2) - IC _{w4,t-1}	0.0355	313	0.0918	392	0.2464	418	2.741***	1.960*	1.789*
53 ARX(2) - IC _{t-1}	0.0365	362	0.0964	413	0.2582	434	2.807***	1.936*	1.757*
54 ARX(2) - IC _{w1,t-1} - SA	0.0380	388	0.1030	438	0.2758	454	2.923***	1.974**	1.763*
55 ARX(2) - IC _{w2,t-1} - SA	0.0376	382	0.1018	435	0.2723	451	2.912***	1.963**	1.767*
56 ARX(2) - IC _{w3,t-1} - SA	0.0378	383	0.1012	434	0.2710	448	2.928***	1.995**	1.784*

(Continued on next page)

Table A.5 – continued

	MSE			DM			HLN		
Model	1-Step	Rank	2-Step	Rank	3-Step	Rank	1-St	2-St	3-St
234 ARMAX(2,2) - IC _{w3,t} - SA	0.0393	414	0.0565	195	0.1583	211	2.601***	1.505	1.167
235 ARMAX(2,2) - IC _{w4,t} - SA	0.0410	436	0.0641	218	0.1305	180	2.932***	1.160	1.335
236 ARMAX(2,2) - IC _t - SA	0.0436	447	0.0889	371	0.1846	261	2.951***	1.371	1.232
237 ARMAX(2,2) - IC _{w1,t-1}	0.0398	426	0.0801	298	0.1862	266	2.533***	1.766*	1.331
238 ARMAX(2,2) - IC _{w2,t-1}	0.0385	396	0.0872	360	0.1948	292	2.401***	1.897*	1.554
239 ARMAX(2,2) - IC _{w3,t-1}	0.0323	248	0.0684	236	0.1680	225	2.002**	1.633	1.410
240 ARMAX(2,2) - IC _{w4,t-1}	0.0347	292	0.0712	230	0.1712	230	2.419***	1.835*	1.640
241 ARMAX(2,2) - IC _{t-1}	0.0384	394	0.0749	278	0.1859	265	2.384***	1.653*	1.562
242 ARMAX(2,2) - IC _{w1,t-1} - SA	0.0313	224	0.0727	262	0.1614	217	1.875**	1.516	1.096
243 ARMAX(2,2) - IC _{w2,t-1} - SA	0.0492	488	0.0859	347	0.1873	271	2.621***	1.726*	1.390
244 ARMAX(2,2) - IC _{w3,t-1} - SA	0.0313	223	0.0732	268	0.2097	343	1.779*	1.501	1.329
245 ARMAX(2,2) - IC _{w4,t-1} - SA	0.0396	424	0.0677	230	0.1614	216	2.818***	1.485	1.487
246 ARMAX(2,2) - IC _{t-1} - SA	0.0399	430	0.0832	322	0.1782	243	1.664*	1.300	1.123
247 ARMAX(2,2) - IC _{w1,t-2}	0.0363	354	0.0969	416	0.2148	362	2.496***	1.890*	1.474
248 ARMAX(2,2) - IC _{w2,t-2}	0.0323	250	0.0864	349	0.2017	308	2.961***	1.792*	1.453
249 ARMAX(2,2) - IC _{w3,t-2}	0.0368	367	0.0800	297	0.1906	281	2.534***	1.426	1.392
250 ARMAX(2,2) - IC _{w4,t-2}	0.0350	300	0.0744	272	0.2013	306	2.062***	1.865*	1.523
251 ARMAX(2,2) - IC _{t-2}	0.0395	420	0.0889	372	0.2354	404	2.114**	1.693*	1.060
252 ARMAX(2,2) - IC _{w1,t-2} - SA	0.0356	317	0.0981	421	0.1915	283	3.378**	1.518	1.177
253 ARMAX(2,2) - IC _{w2,t-2} - SA	0.0518	502	0.1066	455	0.2561	430	2.335**	1.817*	1.216
254 ARMAX(2,2) - IC _{w3,t-2} - SA	0.0363	357	0.0884	368	0.2310	398	2.558***	1.545	1.448
255 ARMAX(2,2) - IC _{w4,t-2} - SA	0.0454	464	0.0777	287	0.1599	214	3.100***	1.572	1.448
256 ARMAX(2,2) - IC _{t-2} - SA	0.0518	501	0.1156	475	0.2278	387	1.720**	1.319	1.102
257 ARX(1) - G _{w1,t}	0.0277	156	0.0479	163	0.1064	132	1.720**	0.906	1.037
258 ARX(1) - G _{w2,t}	0.0212	43	0.0478	162	0.1686	227	1.379	0.810	0.592
259 ARX(1) - G _{w3,t}	0.0227	62	0.0325	58	0.0856	82	1.854**	1.443	1.443
260 ARX(1) - G _{w4,t}	0.0206	32	0.0279	33	0.0556	20	1.771*	1.418	1.025
261 ARX(1) - G _t	0.0166	1	0.0157	1	0.0382	4	0.000	0.000	0.308
262 ARX(1) - G _{w1,t} - SA	0.0294	187	0.0503	175	0.1084	142	2.605***	1.247	1.247
263 ARX(1) - G _{w2,t} - SA	0.0241	82	0.0509	180	0.1847	262	1.608	1.485	0.707
264 ARX(1) - G _{w3,t} - SA	0.0270	139	0.0393	102	0.0913	94	1.684*	1.837*	1.531
265 ARX(1) - G _{w4,t} - SA	0.0222	53	0.0291	37	0.0555	19	1.639	1.319	0.938
266 ARX(1) - G _t - SA	0.0188	12	0.0175	5	0.0383	6	0.998	0.700	0.299
267 ARX(1) - IC _{w1,t} - G _{w1,t}	0.0256	108	0.0443	143	0.1089	145	2.057**	1.591	1.048
268 ARX(1) - IC _{w2,t} - G _{w2,t}	0.0201	26	0.0385	95	0.1056	128	1.087	1.285	1.079
269 ARX(1) - IC _{w3,t} - G _{w3,t}	0.0240	79	0.0372	88	0.1150	157	1.669**	1.969**	1.502
270 ARX(1) - IC _{w4,t} - G _{w4,t}	0.0192	15	0.0296	38	0.0735	55	0.778	1.479	1.035
271 ARX(1) - IC _t - G _t	0.0186	11	0.0242	20	0.0680	45	0.709	1.159	0.757
272 ARX(1) - IC _{w1,t} - G _{w1,t} - SA	0.0221	50	0.0282	34	0.1066	134	1.360	1.008	1.503
273 ARX(1) - IC _{w1,t} - G _{w1,t} - SA	0.0266	133	0.0421	125	0.1043	125	2.231**	1.822*	1.257
274 ARX(1) - IC _{w2,t} - G _{w2,t} - SA	0.0206	33	0.0399	107	0.1090	146	1.205	1.364	1.097
275 ARX(1) - IC _{w3,t} - G _{w3,t} - SA	0.0241	83	0.0353	77	0.0745	143	1.672*	1.953*	1.436
276 ARX(1) - IC _{w4,t} - G _{w4,t} - SA	0.0200	24	0.0312	50	0.0745	57	0.924	1.236	0.872
277 ARX(1) - IC _t - G _t - SA	0.0186	10	0.0186	6	0.0531	17	0.852	0.952	0.681
278 ARX(1) - IC _{w1,t} - G _{w1,t} - SA	0.0241	81	0.0297	39	0.1127	153	1.763**	0.951	1.602
279 ARX(1) - G _{w1,t-1}	0.0292	180	0.0491	168	0.1130	154	1.736**	1.001	1.140
280 ARX(1) - G _{w2,t-1}	0.0313	226	0.0878	364	0.3210	477	1.189	0.661	0.538
281 ARX(1) - G _{w3,t-1}	0.0230	70	0.0314	52	0.0887	77	1.891*	1.287	1.080
282 ARX(1) - G _{w4,t-1}	0.0231	71	0.0354	79	0.0778	67	2.692***	1.813*	1.482
283 ARX(1) - G _{t-1}	0.0182	8	0.0208	9	0.0513	15	1.516	1.514	1.217
284 ARX(1) - G _{w1,t-1} - SA	0.0310	215	0.0530	186	0.1198	167	2.482**	1.151	1.375
285 ARX(1) - G _{w2,t-1} - SA	0.0392	410	0.0977	419	0.3762	504	1.181	0.637	0.521
286 ARX(1) - G _{w3,t-1} - SA	0.0276	151	0.0396	106	0.1018	117	2.057**	1.188	0.993
287 ARX(1) - G _{w4,t-1} - SA	0.0254	105	0.0381	94	0.0802	71	2.291**	1.657**	1.360
288 ARX(1) - G _{t-1} - SA	0.0197	20	0.0220	11	0.0510	14	1.647*	1.402	1.142
289 ARX(1) - IC _{w1,t-1} - G _{w1,t-1}	0.0336	270	0.0779	241	0.1779	241	2.393**	1.632	0.979
290 ARX(1) - IC _{w2,t-1} - G _{w2,t-1}	0.0255	106	0.0581	203	0.1763	236	1.921*	1.334	0.951
291 ARX(1) - IC _{w3,t-1} - G _{w3,t-1}	0.0239	77	0.0312	49	0.0924	95	2.189**	1.251	1.145
292 ARX(1) - IC _{w4,t-1} - G _{w4,t-1}	0.0227	63	0.0344	73	0.0850	80	2.182**	1.657**	1.751*

(Continued on next page)

Table A.5 – continued

Model	MSE				DM			HLN			
	1-Step	Rank	2-Step	Rank	3-Step	Rank	3-Step	Rank	1-St	2-St	3-St
470	0.0257	110	0.0277	32	0.1006	113	1.348	0.768	2.028**	1.194	1.329
471	0.0268	134	0.0569	196	0.1028	120	2.297**	1.876*	3.119***	1.479	1.527
472	0.0379	384	0.0830	319	0.1818	252	2.228**	1.190	1.991**	1.081	1.092
473	0.0299	195	0.0440	140	0.1327	184	1.777*	0.908	2.114**	1.224	0.819
474	0.0445	453	0.0864	350	0.1605	215	1.906*	0.950	1.630	0.849	0.721
475	0.0362	350	0.0577	201	0.0855	81	2.302**	0.861	1.849*	0.787	0.603
476	0.0310	214	0.0799	295	0.1950	293	2.142**	1.260	1.782*	1.117	1.051
477	0.0248	95	0.0406	115	0.0952	105	2.211**	1.770*	3.721***	1.830*	1.689*
478	0.0890	521	0.1227	481	0.5435	521	1.204	0.832	1.218	0.744	0.513
479	0.0226	61	0.0491	169	0.1306	181	1.513	2.441**	4.997***	2.705***	2.462**
480	0.0257	111	0.0441	141	0.0896	91	1.935*	1.821*	3.422***	1.817*	1.879*
481	0.0208	35	0.0350	75	0.0720	52	1.219	1.631	2.677***	1.732*	1.490
482	0.0270	137	0.0428	129	0.0927	97	2.089***	1.270**	4.524***	2.047**	1.646*
483	0.1080	524	0.1657	518	0.6923	526	1.266	0.744	0.668	0.668	0.516
484	0.0203	28	0.0371	87	0.1249	171	0.933	1.434	3.145***	1.609	1.502
485	0.0274	146	0.0376	89	0.0985	110	2.394**	1.519	3.655***	1.659*	1.790*
486	0.0224	55	0.0242	19	0.0685	47	1.757*	1.188	3.114***	1.441	1.299
487	0.0357	325	0.0513	183	0.1330	185	1.661*	2.198**	2.122**	1.739*	1.872*
488	0.0618	510	0.0831	320	0.3927	511	1.119	0.853	1.164	0.772	0.515
489	0.0242	85	0.0475	160	0.1260	172	1.714*	1.876*	3.482***	1.860*	1.809*
490	0.0260	119	0.0425	127	0.1009	115	1.865*	1.968**	2.639***	1.850*	1.749*
491	0.0177	4	0.0273	29	0.0620	34	0.328	1.399	1.912*	1.534	1.736*
492	0.0456	468	0.0603	429	0.2532	429	1.633	1.002	1.512	0.849	0.656
493	0.0413	439	0.0647	200	0.1442	199	1.875*	1.817*	2.341**	1.540	1.425
494	0.0437	448	0.0847	335	0.2450	416	2.599***	1.416	2.383**	1.229	0.762
495	0.0222	51	0.0400	109	0.1188	163	1.266	1.479	3.487***	1.512	1.248
496	0.0276	154	0.0386	96	0.0993	112	2.277**	1.632	3.758***	1.701*	1.920*
497	0.0301	199	0.0407	116	0.0975	140	1.788*	1.400	1.324**	1.288	0.982
498	0.0629	511	0.0986	425	0.4177	515	1.846*	1.263	1.766*	1.130	0.744
499	0.0210	39	0.0366	85	0.0767	62	1.241	1.522	2.889***	1.420	1.585
500	0.2315	527	0.3639	528	0.6376	524	1.078	0.578	1.053	0.545	0.482
501	0.0229	66	0.0428	131	0.1205	168	1.633	2.359**	2.142**	2.508**	2.210**
502	0.0258	114	0.0428	130	0.0905	92	2.445**	1.715*	3.201***	1.602	1.884*
503	0.0235	74	0.0341	69	0.0803	72	1.788*	1.473	2.687***	1.418	1.194
504	0.0263	127	0.0378	90	0.0886	86	2.055**	1.728*	3.505***	1.758*	1.612
505	0.1147	525	0.1565	511	0.5446	522	1.382	0.710	1.432	0.651	0.715
506	0.0250	98	0.0393	104	0.1218	169	1.956*	1.784*	3.842***	1.718*	1.430
507	0.0352	303	0.0477	161	0.1195	166	3.313***	1.313	4.311***	1.306	1.540
508	0.0244	88	0.0360	82	0.0774	66	1.658*	1.486	3.019***	1.476	1.462
509	0.0302	202	0.0652	221	0.1355	192	2.045**	1.732*	2.573**	1.540	1.142
510	0.0539	504	0.0847	333	0.2840	460	2.002**	1.676*	1.947*	1.610	0.874
511	0.0276	150	0.0560	193	0.1540	208	2.723***	1.901*	4.243***	2.246**	1.831*
512	0.0291	179	0.0435	137	0.1029	121	2.105**	1.647**	3.218***	1.545	1.580
513	0.0307	211	0.0432	134	0.0942	100	4.083***	1.970**	4.064***	1.822*	1.694*
514	0.0761	518	0.2362	524	0.7232	528	3.285***	1.705*	2.877***	1.499	1.221
515	0.0278	157	0.0672	229	0.1646	221	2.076**	1.581	3.698***	1.720*	1.153
516	0.0594	509	0.0969	415	0.2972	470	2.810***	1.625	3.520***	1.386	1.015
517	0.0332	266	0.0744	273	0.2065	329	2.754***	1.749*	3.530***	1.529	1.172
518	0.0552	505	0.0667	228	0.1718	231	2.995***	1.310	3.208***	1.329	1.356
519	0.0363	353	0.0459	152	0.0972	108	2.973***	1.756*	2.718***	1.796*	1.423
520	0.1010	523	0.2831	526	0.8950	529	3.195***	1.869*	3.289***	1.740*	1.442
Nonlinear models											
521	0.0332	262	0.0388	101	0.0589	28	2.434**	1.053	2.925***	1.720*	1.447
522	0.0368	366	0.0447	144	0.062	35	2.497**	1.190	3.015***	1.779*	1.411
523	0.0342	280	0.0436	138	0.0652	40	2.337**	1.183	2.903***	1.721*	1.389
State-level models											
524	0.2845	528	0.3391	527	0.3966	513	5.300***	2.770***	4.917***	2.306**	2.306**
525	0.0292	181	0.0310	48	0.0411	7	-0.133	-0.299	2.681***	1.308	1.308
526	0.0299	196	0.0314	51	0.0413	8	-0.062	-0.283	2.746***	1.324	1.324

(Continued on next page)

Table A.5 – continued

Model	MSE						DM			HLN		
	1-Step	Rank	2-Step	Rank	3-Step	Rank	1-St	2-St	3-St	1-St	2-St	3-St
527 IU active × LF	0.0296	190	0.0318	56	0.0423	9	-0.091	-0.264	-1.137	2.686***	1.303	1.303
528 IU unempl. × LF	0.0298	194	0.0322	57	0.0425	10	-0.069	-0.251	-1.133	2.712***	1.312	1.312
529 IU unempl. × unempl.	0.0917	522	0.0690	242	0.0618	33	2.335**	0.648	-0.531	3.334***	1.661*	1.661*

Notes: Full sample: 1967:1-2009:6; short sample: 2004:1-2009:6. In all panels ***, ** and * indicate rejection at 1, 5 and 10%, respectively.

Table A.6: US states and internet diffusion among total population, active population, unemployed population

N.	State	All	Act.	Une.	N.	State	All	Act.	Une.
0	United States	1.000	1.000	1.000	26	Missouri	0.969	0.975	0.865
1	Alabama	0.872	0.904	0.655	27	Montana	1.012	1.023	1.061
2	Alaska	1.141	1.096	1.096	28	Nebraska	1.054	1.035	1.061
3	Arizona	0.981	0.995	1.097	29	Nevada	1.022	0.992	0.804
4	Arkansas	0.869	0.904	0.688	30	New Hampshire	1.143	1.110	1.056
5	California	1.000	0.998	1.008	31	New Jersey	1.045	1.022	0.935
6	Colorado	1.061	1.035	0.925	32	New Mexico	0.978	0.980	1.215
7	Connecticut	1.058	1.046	0.860	33	New York	0.978	0.988	1.099
8	Delaware	1.018	1.008	1.010	34	North Carolina	0.954	0.959	0.992
9	D. of Columbia	1.032	1.039	0.908	35	North Dakota	1.054	1.050	0.880
10	Florida	0.979	0.966	0.980	36	Ohio	1.002	1.006	1.005
11	Georgia	0.989	0.985	1.074	37	Oklahoma	0.908	0.928	0.905
12	Hawaii	1.035	1.016	1.097	38	Oregon	1.034	1.013	1.109
13	Idaho	0.965	0.948	1.052	39	Pennsylvania	1.007	1.006	1.021
14	Illinois	1.036	1.035	1.014	40	Rhode Island	1.041	1.029	0.948
15	Indiana	0.982	0.993	0.773	41	South Carolina	0.956	0.964	0.853
16	Iowa	1.038	1.012	0.869	42	South Dakota	1.072	1.043	1.055
17	Kansas	1.070	1.045	0.984	43	Tennessee	0.957	0.965	1.144
18	Kentucky	0.953	1.006	1.094	44	Texas	0.935	0.939	0.950
19	Louisiana	0.926	0.955	0.799	45	Utah	1.132	1.087	1.253
20	Maine	1.072	1.087	1.127	46	Vermont	1.107	1.075	1.143
21	Maryland	1.069	1.038	0.939	47	Virginia	1.045	1.015	1.172
22	Massachusetts	1.070	1.078	1.118	48	Washington	1.120	1.104	1.140
23	Michigan	1.014	1.025	1.008	49	West Virginia	0.872	0.941	1.010
24	Minnesota	1.114	1.079	1.156	50	Wisconsin	1.085	1.068	1.054
25	Mississippi	0.867	0.888	0.684	51	Wyoming	1.102	1.067	1.110

Notes: Authors calculations using the October 2007 CPS computer use supplement. State internet diffusion is expressed in relative terms with the the federal average normalized to one. The actual diffusion at the national level is equal to 76.2, 82.6 and 76.5 respectively for total, active and unemployed population.

Table A.7: Forecasting US unemployment rate (u_t) in levels. Best 15 models in terms of lowest MSE, best models without GI and non-linear models.

1-step ahead				2-step ahead				3-step ahead			
n. Model	MSE Rank	DM	HLN	n. Model	MSE Rank	DM	HLN	n. Model	MSE Rank	DM	HLN
Panel A1: Best models											
403 ARMAX(1,1) - IC _t - G _t	0.0167	1	-	332 ARX(2) - G _t - SA	0.0169	1	-	Panel A3: Best models			
393 ARMAX(1,1) - G _t	0.0183	2	0.927	327 ARX(2) - G _t	0.0184	2	0.487	334 ARX(2) - G _{t-1} - SA	0.0482	1	-
327 ARX(2) - G _t	0.0186	3	0.676	459 ARMAX(2,2) - G _t	0.0214	3	0.500	354 ARX(2) - G _{t-1} - SA	0.0518	2	0.280
425 ARMAX(1,1) - IC _{t-1} - G _{t-1}	0.0187	4	1.147	349 ARX(2) - G _{t-1}	0.0215	4	1.456	327 ARX(2) - G _t	0.0529	3	0.386
459 ARMAX(2,2) - G _t	0.0189	5	1.155	371 ARX(2) - G _{t-2}	0.0218	5	1.559	266 ARX(1) - G _t - SA	0.0535	4	0.226
332 ARX(2) - G _t - SA	0.0191	6	1.097	491 ARMAX(2,2) - IC _{t-1} - G _{t-1}	0.0228	6	0.950	459 ARMAX(2,2) - G _t	0.0547	5	0.356
371 ARX(2) - G _{t-2}	0.0192	7	0.786	403 ARMAX(1,1) - IC _t - G _t	0.0233	7	0.697	491 ARMAX(2,2) - IC _{t-1} - G _{t-1}	0.0554	6	0.407
437 ARMAX(1,1) - G _{t-2}	0.0193	8	0.819	354 ARX(2) - G _{t-1} - SA	0.0237	8	1.087	261 ARX(1) - G _t	0.0569	7	0.357
481 ARMAX(2,2) - G _{t-1}	0.0194	9	1.450	343 ARX(2) - IC _t - G _t - SA	0.0240	9	1.483	349 ARX(2) - G _{t-1}	0.0596	8	1.232
343 ARX(2) - IC _t - G _t - SA	0.0197	10	1.037	359 ARX(2) - IC _{t-1} - G _{t-1}	0.0244	10	1.894*	376 ARX(2) - G _{t-2} - SA	0.0599	9	0.827
469 ARMAX(2,2) - IC _t - G _t	0.0197	11	1.820*	393 ARMAX(1,1) - G _t	0.0246	11	0.750	403 ARMAX(1,1) - IC _t - G _t	0.0601	10	0.561
415 ARMAX(1,1) - G _{t-1}	0.0197	12	1.716*	469 ARMAX(2,2) - IC _t - G _t	0.0248	12	0.805	393 ARMAX(1,1) - G _t	0.0615	11	0.632
491 ARMAX(2,2) - IC _{t-1} - G _{t-1}	0.0197	13	1.332	365 ARX(2) - IC _{t-1} - G _{t-1} - SA	0.0252	13	1.307	365 ARX(2) - IC _{t-1} - G _{t-1} - SA	0.0618	12	1.099
409 ARMAX(1,1) - IC _t - G _t - SA	0.0200	14	1.251	376 ARX(2) - G _{t-2} - SA	0.0253	14	1.260	425 ARMAX(1,1) - IC _{t-1} - G _{t-1}	0.0624	13	0.840
420 ARMAX(1,1) - G _{t-1} - SA	0.0200	15	1.172	261 ARX(1) - G _t	0.0253	15	0.941	481 ARMAX(2,2) - G _{t-1}	0.0626	14	0.718
								398 ARMAX(1,1) - G _t - SA	0.0630	15	0.865
Panel B1: Best models without Google											
127 ARMAX(2,2) - IC _{w4,t-2} - SA	0.0269	97	1.925*	122 ARMAX(2,2) - IC _{w4,t-2}	0.0581	170	1.927*	Panel B3: Best models without Google			
205 ARMAX(1,1) - IC _{w4,t} - SA	0.0303	172	1.969**	160 ARX(1) - IC _{w4,t-2}	0.0694	208	2.038**	122 ARMAX(2,2) - IC _{w4,t-2}	0.1549	174	1.548
								134 ARMAX(1,1) - SA	0.1787	205	1.264
Panel C1: Non-linear models											
521 SETAR(2)	0.0511	491	2.967***	521 SETAR(2)	0.1750	509	2.087**	Panel C3: Non-linear models			
522 LSTAR(2)	0.0518	493	3.001***	522 LSTAR(2)	0.1746	508	2.080**	521 SETAR(2)	0.4154	502	1.701*
523 AAR(2)	0.0554	498	3.111***	523 AAR(2)	0.1851	510	1.972**	522 LSTAR(2)	0.4156	503	1.667*
								523 AAR(2)	0.4341	505	1.609

Notes: ***, ** and * indicate rejection at 1, 5 and 10%, respectively. This table reports the best 15 models in terms of MSE among the 523 estimated ones. The complete list of models and their forecasting performance is available in the Appendix (table A.5). SA indicates the model augmented with a multiplicative seasonal factor.

Table A.8: Forecasting US unemployment rate in logs ($\log(u_t)$). Best 15 models, best models without GI and non-linear models.

1-step ahead				2-step ahead				3-step ahead						
n. Model	MSE	Rank	DM	HLN	n. Model	MSE	Rank	DM	HLN	n. Model	MSE	Rank	DM	HLN
Panel A1: Best models														
327 ARX(2) - G_t	0.0191	1	-	-	327 ARX(2) - G_t	0.0237	1	-	-	266 ARX(1) - $G_t - SA$	0.0503	1	-	-
337 ARX(2) - $IC_t - G_t$	0.0224	2	1.700*	1.882*	261 ARX(1) - G_t	0.0270	2	0.248	0.877	261 ARX(1) - G_t	0.0543	2	0.422	0.632
398 ARMAX(1,1) - $G_t - SA$	0.0224	3	1.026	2.406**	332 ARX(2) - $G_t - SA$	0.0276	3	0.563	0.680	288 ARX(1) - $G_{t-1} - SA$	0.0627	3	1.028	0.994
425 ARMAX(1,1) - $IC_{t-1} - G_{t-1}$	0.0232	4	0.988	1.847*	343 ARX(2) - $IC_t - G_t - SA$	0.0278	4	0.687	0.950	310 ARX(1) - $G_{t-2} - SA$	0.0679	4	1.527	1.719*
326 ARX(2) - $G_{w4,t}$	0.0234	5	1.186	1.734*	265 ARX(1) - $G_{w4,t} - SA$	0.0334	5	0.655	1.684*	283 ARX(1) - G_{t-1}	0.0732	5	1.243	1.313
331 ARX(2) - $G_{w4,t} - SA$	0.0236	6	1.449	1.957*	349 ARX(2) - G_{t-1}	0.0347	6	1.170	0.845	332 ARX(2) - $G_t - SA$	0.0750	6	0.780	1.095
338 ARX(2) - $IC_{w1,t} \dots IC_{w4,t}$	0.0238	7	1.524	3.559***	260 ARX(1) - $G_{w4,t}$	0.0353	7	0.650	1.462	265 ARX(1) - $G_{w4,t} - SA$	0.0763	7	0.563	0.819
332 ARX(2) - $G_t - SA$	0.0241	8	0.954	1.061	371 ARX(2) - G_{t-2}	0.0376	8	1.336	0.933	305 ARX(1) - G_{t-2}	0.0789	8	1.481	1.279
336 ARX(2) - $IC_{w4,t} - G_{w4,t}$	0.0244	9	1.406	1.806*	331 ARX(2) - $G_{w4,t} - SA$	0.0377	9	0.605	0.814	354 ARX(2) - $G_{t-1} - SA$	0.0804	9	0.994	1.202
469 ARMAX(2,2) - $IC_t - G_t$	0.0246	10	1.701*	2.181**	359 ARX(2) - $IC_{t-1} - G_{t-1}$	0.0379	10	1.317	0.994	327 ARX(2) - G_t	0.0828	10	0.848	1.023
349 ARX(2) - G_{t-1}	0.0248	11	1.081	1.130	266 ARX(1) - $G_t - SA$	0.0380	11	0.451	0.633	260 ARX(1) - $G_{w4,t}$	0.0841	11	0.575	0.776
371 ARX(2) - G_{t-2}	0.0248	12	2.025**	2.492**	338 ARX(2) - $IC_{w1,t} \dots IC_{w4,t}$	0.0382	12	0.822	1.341	376 ARX(2) - $G_{t-2} - SA$	0.0908	12	1.108	1.300
343 ARX(2) - $IC_t - G_t - SA$	0.0248	13	1.490	1.740*	326 ARX(2) - $G_{w4,t}$	0.0398	13	0.665	0.848	408 ARMAX(1,1) - $IC_{w4,t} - G_{w4,t} - SA$	0.0911	13	1.043	1.281
475 ARMAX(2,2) - $IC_t - G_t - SA$	0.0249	14	1.637	2.711***	337 ARX(2) - $IC_t - G_t$	0.0404	14	1.462	1.336	365 ARX(2) - $IC_{t-1} - G_{t-1} - SA$	0.0932	14	1.170	1.248
260 ARX(1) - $G_{w4,t}$	0.0252	15	1.421	2.794***	381 ARX(2) - $IC_{t-2} - G_{t-2}$	0.0431	15	1.246	0.915	349 ARX(2) - G_{t-1}	0.0989	15	1.293	1.306
Panel B1: Best models without Google														
127 ARMAX(2,2) - $IC_{w4,t-2} - SA$	0.0255	17	1.560	2.848***	Panel B2: Best models without Google					122 ARMAX(2,2) - $IC_{w4,t-2}$	0.1348	37	1.650*	1.417
129 AR(1)	0.0430	258	2.203**	2.893***	129 AR(1)					0.1018	228	1.599	1.658*	1.120
Panel C1: Non-linear models														
521 SETAR(2)	0.0513	308	2.768***	3.556***	Panel C2: Non-linear models					521 SETAR(2)	0.4173	357	1.758*	1.529
522 LSTAR(2)	0.0514	309	2.759***	3.509***	522 LSTAR(2)					0.1770	371	2.130**	1.971**	1.537
523 AAR(2)	0.0616	390	3.023***	3.591***	523 AAR(2)					0.2083	434	1.970**	1.832*	1.452

Notes: ***, ** and * indicate rejection at 1, 5 and 10%, respectively. This table reports the best 15 models in terms of MSE among the 523 estimated ones. The complete list of models and their forecasting performance is available in the Appendix (table A.5). SA indicates the model augmented with a multiplicative seasonal factor.