Macroeconomic Nowcasting Using Google Probabilities*

Gary Koop, University of Strathclyde†

Luca Onorante, European Central Bank‡

August 2013

Abstract: Many recent papers have investigated whether data from internet search engines such as Google can help improve nowcasts or short-term forecasts of macroeconomic variables. These papers construct variables based on Google searches and use them as explanatory variables in regression models. We add to this literature by nowcasting using dynamic model selection (DMS) methods which allow for model switching between time-varying parameter regression models. This is potentially useful in an environment of coefficient instability and over-parameterization which can arise when forecasting with Google variables. We extend the DMS methodology by allowing for the model switching to be controlled by the Google variables through what we call Google probabilities. That is, instead of using Google variables as regressors, we allow them to determine which nowcasting model should be used at each point in time. In an empirical exercise involving nine major monthly US macroeconomic variables, we find DMS methods to provide large improvements in nowcasting. Our use of Google model probabilities within DMS often performs better than conventional DMS.

Keywords: internet search data, nowcasting, dynamic model averaging, state space model

*This working paper should not be reported as representing the views of the ECB. The views expressed are those of the authors and do not necessarily reflect those of the ECB.

†This research was supported by the ESRC under grant RES-062-23-2646. Gary Koop is a Fellow at the Rimini Centre for Economic Analysis. Address for correspondence: Gary Koop, Department of Economics, University of Strathclyde, 130 Rottenrow, Glasgow G4 0GE, UK. Email: Gary.Koop@strath.ac.uk

‡Currently on leave to the Central Bank of Ireland

1 Introduction

Macroeconomic data are typically published with a time lag. This has led to a growing body of research on nowcasting. Nowcasting uses currently available data to provide timely estimates of macroeconomic variables weeks or even months before their initial estimates are produced. The availability of internet search data has provided a new resource for researchers interested in nowcasts or short-term forecasts of macroeconomic variables. Google search data, available since January 2004, is a particularly popular source. Pioneering papers such as Choi and Varian (2009, 2011) have led to an explosion of nowcasting work using Google data including, among many others, Artola and Galan (2012), Askitas and Zimmermann (2009), Carriere-Swallow and Labbe (2011), Chamberlin (2010), D'Amuri and Marcucci (2009), Hellerstein and Middeldorp (2012), Kholodilin, Podstawski and Siliverstovs (2010), McLaren and Shanbhoge (2011), Scott and Varian (2012), Schmidt and Vosen (2009), Suhoy (2009) and Wu and Brynjolfsson (2010).[1]

These papers report a variety of findings for a range of variables, but a few general themes emerge. First, Google data is potentially useful in nowcasting or short-term forecasting, but there is little evidence that it can be successfully used for long-term forecasting. Second, Google data is only rarely found to be useful for broad macroeconomic variables (e.g. inflation, industrial production, etc.)[2] and is more commonly used to nowcast specific variables relating to consumption, housing or labor markets. For instance, Choi and Varian (2011) successfully nowcast the variables motor vehicles and car parts, initial claims for unemployment benefits and tourist arrivals in Hong Kong. Third, the existing literature uses linear regression methods.

The present paper deals with the second and third of these points. We nowcast a variety of conventional US monthly macroeconomic variables and see if Google variables provide additional nowcasting power beyond a conventional set of predictors. It is common (see, among many others, Giannone, Lenza, Momferatou and Onorante, 2010) to forecast inflation using a variety of macro predictors such as unemployment, the term spread, wage inflation, oil price inflation, etc.. We use Google variables in different ways as additional information and check whether their inclusion can improve nowcasting power. We do this for nine different macroeconomic variables.

The main innovations in our approach relate to the manner in which we include the Google variables in our regression models. The first is that we use Dynamic model averaging and model selection (DMA and DMS) methods with time-varying parameter (TVP) regressions. DMA methods for TVP regression models were developed in Raftery et al (2010) and have been used them successfully in several applications (e.g., among others, Dangl and Halling, 2012, Koop and Korobilis, 2012, Koop and Onorante, 2012, Koop and Tole, 2013, Nicoletti

[1] This list of papers uses Google data for macroeconomic forecasting. Google data is also being used for nowcasting in other fields such as finance and epidemiology.

[2] A notable exception is the nowcasting of U.S. unemployment in D'Amuri and Marcucci (2009).

and Passaro, 2012).

Initially we implement DMA and DMS in a conventional manner, using Google variables as additional predictors in TVP regressions. This represents a useful extension over existing nowcasting methods, such as Choi and Varian (2009, 2011), who use linear regression methods with constant coefficients. The second innovative aspect of the paper is that we extend the DMA methodology to use the Google data in a different manner. Instead of simply using a Google variable as an explanatory variable in a regression, we develop a method which allows for the inclusion probability of each macro explanatory variable to depend on the Google data. This motivates the terminology used in the title of this paper: "Google probabilities". The rational behind our approach is that some of the existing literature (e.g. Choi and Varian, 2011) suggests that Google variables might not be good linear predictors. However, they may be good at signalling turning points or other forms of change or model switching. In particular, we hypothesize that Google searches are able to collect "collective wisdom"and be informative about which macro variables are important in the model at different points in time, either directly or by influencing the outcomes through expectations. For example, a surge in searches about oil prices may not say much per se on whether oil prices are increasing or decreasing, but may indicate that the variable should be relevant in modelling. This should trigger a switch towards nowcasting models including the oil price as explanatory variable.

In an empirical exercise involving monthly US data on nine macroeconomic variables, we find DMS methods to nowcast well, regardless of whether they involve Google model probabilities or not. In particular, DMS tends to nowcast slightly better than DMA and much better than standard benchmarks using OLS methods. The use of Google probabilities to influence model switching often leads to further improvements in nowcast performance.

2 Macroeconomic Nowcasting and Google Data

Table 1 lists the macroeconomic variables we are interested in nowcasting. We use monthly US data from January 1973 through July 2012. Note that, as is commonly done, all of our variables are transformed so as to be rates (e.g. inflation rate, unemployment rate, etc.). All data are taken from the BIS Macroeconomic series databases, OECD Main Economic Indicators (OECD), Hamburg World Economic Archive and the Federal Reserve Bank of Chicago.

Table 1: Dependent and Explanatory Variables	
Variable	Raw Variable (w_t)
Inflation	Consumer price index, all items
Wage inflation	Ave. hourly earnings in manuf.
Unemployment	Unemployment rate, all employees
Term spread	Long minus short 10 yr. Treasury minus Fed funds rate
FCI	Financial Conditions IndexWe use the one produced by the Chicago Fed who ca
Comm. price inflation	Price Index, food and energy
Industrial production	Total industrial production excluding construction
Oil price inflation	Crude oil price (USD per barrel)
Money supply growth	Money supply (M3)

We use the one produced by the Chicago Fed who calibrate this variable to have an average value of zero and a standard deviation of one. Positive/negative numbers indicate tighter/looser than average financial conditions.

Corresponding to each of these variables, we produce a composite Google search variable. Of course, for any concept there are many potential Google search terms and there are different treatments of this in the literature. For example, Scott and Varian (2012) use 151 search categories.[3] In this paper we use a standardized procedure with the scope of minimizing the amount of judgement in the choice of variables. We start by searching for the name of the macro variable of interest and we collect the corresponding Google search volume. Along with this variable, the Google interface supplies a set of related terms. These are the most popular terms related to the search: Google chooses them in a mechanical manner, by examining searches conducted by users immediately before and after. We fetch these related searches, and we repeat the procedure for each of them, finding new terms. Only at this point some judgment is necessary. The related searches in Google are found automatically, therefore terms completely unrelated to economic concepts are removed manually. We could alternatively have chosen to limit our search to some specific Google category, but those are also defined automatically and remaining extraneous variables would have needed manual intervention. It is important, however, to note that

[3]Categories are aggregates of searches that are classified by the Google engine as belonging to a specific category. Examples of top-level categories are 'Food and beverages' or 'News and current events'. Running a regression with 151 explanatory categories, using data beginning in January 2004, is a challenge, raising concerns about over-fitting. They address these problems by using Bayesian variable selection methods, involving a spike-and-slab prior, to obtain a more parsimonious model. Their work well illustrates the two problems which must be addressed with Google data: i) how to select the Google search variables and ii) given the number of Google search variables is typically large, how to ensure parsimony.

variables are not eliminated on the basis of (expected) performance, but only when they are obvious mistakes (e.g. when searching for “spread” all results related to food are not retained). We also mechanically deleted all repeated terms, a frequent event when using the concept of “related” more than once. The remaining Google variables are attributed to the macro variable used to start the search.

Our final Google database is composed of 259 search results (see the Appendix for a complete list). All series start at the beginning of 2004 and each volume search is separately normalized from 0 to 100. Variables searched with high volume have weekly frequency; less searched terms are supplied by the Google interface as monthly observations. Our research and the data to be forecasted are at most at monthly frequency, therefore we convert the weekly series by taking the last observation available for every month.

Thus, for each of the 9 macroeconomic variables in Table 1, we match a number of Google search variables. For each variable, we have, on average, over 20 Google search variables, unevenly distributed. To ensure parsimony, we adopt a strategy of averaging all the Google search variables to produce a single “Google variable” corresponding to each macroeconomic variable. Such a strategy works well, although other more sophisticated methods (e.g. using principal components methods) would be possible.[4]

The end result is a data set involving 18 variables. These are the 9 variables listed in Table 1 and, corresponding to each, the average Google search variable reflecting internet search activity relating to the underlying macroeconomic concept.

3 Models

Each of our models involves using one of the macroeconomic variables as a dependent variable, y_t, with the remainder of the macroeconomic variables being included as potential explanatory variables, X_t. The Google variables corresponding to X_t will be labelled Z_t. The Google variables are available weekly, whereas the macroeconomic variables are available monthly. In our empirical work, we use the Google data from the last week of month t and, thus, Z_t is data which will be available at the end of the last week of month t. Of course, other timing conventions are possible depending on when nowcasts are desired. For instance, if we used first week of the month Google data we could provide nowcasts after this first week.

3.1 Our Baseline: Regressions with Constant Coefficients

A standard, one-step ahead regression model for forecasting y_t (e.g., among many others, Koop and Korobilis, 2012) would be:

[4]Note that the macroeconomic variables and Google variables have different time spans since the internet search data is not available before January 2004. We will discuss how we treat this issue in a subsequent section.

$$y_t = X'_{t-1}\beta + \varepsilon_t. \tag{1}$$

Typically, the model would also include lags of the dependent variable and an intercept. All models and all the empirical results in this paper include these (with a lag length of 2), but for notational simplicity we will not explicitly note this in the formulae in this section.

We then add the Google regressors. We assume the following timing convention: At the end of month t or early in month $t+1$, we assume y_t has not been observed and, hence, we are interested in nowcasts of it. The Google search data for the last week of month t, Z_t, becomes available. The other macroeconomic variables are released with a time lag so that X_{t-1} is available, but not X_t. With these assumptions about timing, the following regression can be used for nowcasting y_t early in month $t+1$

$$y_t = X'_{t-1}\beta + Z'_t\gamma + \varepsilon_t. \tag{2}$$

The results in this paper adopt this timing convention, but other timing conventions (e.g. nowcasting in the middle of a month) can be accommodated with minor alterations of the preceding equation (depending on the release date of the variables in X_t).

3.2 TVP Regression Models, Model Averaging and Model Switching with Google regressors

The regressions in the preceding sub-section have two potential problems: i) they assume coefficients are constant over time which, for many macroeconomic time series, is rejected by the data (see, among many other, Stock and Watson, 1996) and ii) they may be over-parameterized since the regressions potentially have many explanatory variables and the time span of the data may be short.

An obvious way to surmount the first problem is to use a TVP regression model. TVP regression models (or multivariate extensions) are increasingly popular in macroeconomics (see, among many others, Canova, 1993, Cogley and Sargent, 2001, 2005, Primiceri, 2005, Canova and Gambetti, 2009, Canova and Ciccarelli, 2009, and Koop, Leon-Gonzalez and Strachan, 2009, and Chan et al, 2012). Our TVP regression model is specified as:

$$\begin{aligned} y_t &= W'_t\theta_t + \varepsilon_t \\ \theta_{t+1} &= \theta_t + \eta_t, \end{aligned} \tag{3}$$

where, in our empirical work, we consider both $W_t = X_{t-1}$ and $W_t = [X_{t-1}\prime, Z'_t]'$. Note that W_t defined in this way includes all information available for nowcasting y_t at the end of month t. We also wish to allow for time variation in the error variance and, thus, ε_t is assumed to be i.i.d. $N\left(0, \sigma_t^2\right)$, where σ_t^2 is replaced by an Exponentially Weighted Moving Average (EWMA) estimate (see RiskMetrics, 1996 and West and Harrison, 1997 and note that EWMA is a special case of a GARCH model):

$$\widehat{\sigma}_t = \kappa \widehat{\sigma}_{t-1} + (1-\kappa)\,\widehat{\varepsilon}_t \widehat{\varepsilon}_t', \tag{4}$$

where $\widehat{\varepsilon}_t$ are the estimated regression errors. We set the decay factor, $\kappa = 0.96$ following suggestions in Riskmetrics (1996). Furthermore, η_t are independent $N(0, Q_t)$ random variables (also independent of ε_t). An advantage of such models is that they are state space models and, thus, standard methods for estimating them exist (e.g. involving the Kalman filter). However, a possible disadvantage is they can be over-parameterized, exacerbating the second problem noted above.

Due to over-parametrization concerns, there is a growing literature which uses model averaging or selection methods in TVP regressions. That is, instead of working with one large over-parameterized model, parsimony can be achieved by averaging over (or selecting between) smaller models. Thus, model averaging or model selection methods can be used to ensure shrinkage in over-parameterized models. With TVP models, it is often desirable to do this in a time-varying fashion and, thus, DMA or DMS methods can be used (see, e.g., Koop and Korobilis, 2012). These allow for a different model to be selected at each point in time (with DMS) or different weighs used in model averaging at each point in time (with DMA). For instance, in light of Choi and Varian (2011)'s finding that Google variables predict better at some points in time than others, one may wish to include the Google variables at some times but not others. DMS allows for this. It can switch between models which include Google variables and models which do not, as necessary.

The pioneering paper which developed methods for DMA and DMS was Raftery et al (2010). Since this paper describes (and provides motivation for) the DMA algorithm used in this paper, we will not provide complete details here. Instead we just describe the model space under consideration and the general ideas involved in the algorithm.

Instead of working with the single regression of the form (**??**), we have $j = 1, .., J$ TVP regression models, each of the form:

$$\begin{aligned} y_t &= W_t^{(j)} \theta_t^{(j)} + \varepsilon_t^{(j)} \\ \theta_{t+1}^{(j)} &= \theta_t^{(j)} + \eta_t^{(j)}, \end{aligned} \tag{5}$$

where $\varepsilon_t^{(j)}$ is $N\left(0, \sigma_t^{2(j)}\right)$ and $\eta_t^{(j)}$ is $N\left(0, Q_t^{(j)}\right)$. The $W_t^{(j)}$ contain different sub-sets of the complete set W_t of potential explanatory variables. If we denote S as the number of explanatory variables in W_t (e.g. in TVP regressions which do not include Google variables, then $S = 8$ since one of the macroeconomic variables in Table 1 will be the dependent variable and the remaining 8 will enter in lagged form as explanatory variables), then there are $J = 2^S$ possible TVP regressions involving every possible combination of the S explanatory variables. Unless S is small, it can be seen that the model space is huge. As discussed in Koop and Korobilis (2012), exact Bayesian estimation of this many TVP regression models using Markov Chain Monte Carlo (MCMC) is computationally infeasible which motivates our use of EWMA and forgetting factor methods.

Within a single TVP regression model we estimate $\sigma_t^{2(j)}$ using EWMA methods (as described above) and $Q_t^{(j)}$ using forgetting factor methods. Forgetting factors have long been used in the state space literature to simplify estimation. Sources such as Raftery et al (2010) and West and Harrison (1997) describe forgetting factor estimation of state space models and we will not repeat this material here. Suffice it to note that they involve choice of a scalar forgetting factor $\lambda \in [0,1]$ and lead to estimates of $\theta_t^{(j)}$ where observations j periods in the past have weight λ^j. An alternative way of interpreting λ is to note that it implies an effective window size of $\frac{1}{1-\lambda}$. With EWMA and forgetting factor methods used to estimate $\sigma_t^{2(j)}$ and $Q_t^{(j)}$, all that is required is the use of the Kalman filter in order to provide estimates of the states and, crucially for our purposes, the predictive density, $p_j\left(y_t|W_{1:t}, y_{1:t-1}\right)$, where $W_{1:t} = (W_1, .., W_t)$ and $y_{1:t-1} = (y_1, .., y_{t-1})$.

DMA and DMS involve a recursive updating scheme using quantities which we label $q_{t|t,j}$ and $q_{t|t-1,j}$. The latter is the key quantity: it is the probability that model j is the model used for nowcasting y_t, at time t, using data available at time $t-1$. The former updates $q_{t|t-1,j}$ using data available at time t. DMS involves selecting the single model with the highest value for $q_{t|t-1,j}$ and using it for forecasting y_t. Note that DMS allows for model switching: at each point in time it is possible that a different model is used for forecasting. DMA uses forecasts which average over all $j = 1, .., J$ models using $q_{t|t-1,j}$ as weights. Note that DMA is dynamic since these weights can vary over time.

Raftery et al (2010) derive the following model updating equation:

$$q_{t|t,j} = \frac{q_{t|t-1,j}\, p_j\left(y_t|W_{1:t}, y_{1:t-1}\right)}{\sum_{l=1}^{J} q_{t|t-1,l}\, p_l\left(y_t|W_{1:t}, y_{1:t-1}\right)} \tag{6}$$

where $p_j\left(y_t|W_{1:t}, y_{1:t-1}\right)$ is the predictive likelihood (i.e. the predictive density for y_t produced by the Kalman filter run for model j evaluated at the realized value for y_t). The algorithm then uses a forgetting factor, α, set to 0.99 following Raftery et al (2010), to produce a model prediction equation:

$$q_{t|t-1,j} = \frac{q_{t-1|t-1,j}^{\alpha}}{\sum_{l=1}^{J} q_{t-1|t-1,l}^{\alpha}}. \tag{7}$$

Thus, starting with $q_{0|0,j}$ (for which we use the noninformative choice of $q_{0|0,j} = \frac{1}{J}$ for $j = 1, .., J$) we can recursively calculate the key elements of DMA: $q_{t|t,j}$ and $q_{t|t-1,j}$ for $j = 1, .., J$.

3.3 DMA and DMS with Google Model Probabilities

Our third and most original contribution consists of using the Google variables not directly as regressors, but as providing information to determine which macroeconomic variables should be included at each point in time. The underlying intuition is that the search volume might show the relevance of a certain variable for nowcasting at one point in time rather than a precise and signed cause-effect relationship. Therefore even those Google searches showing little direct forecasting power as explanatory variables in a regression might be useful in selecting the explanatory variables of most use for nowcasting at any given point in time. Motivated by these considerations, we propose to modify the conventional DMA/DMS methodology as follows.

Let $Z_t = (Z_{1t}, .., Z_{kt})'$ be the vector of Google variables and remember that we construct our data set so that each macroeconomic variable is matched up with one Google variable. Z_{it} is standardized by Google to be a number between 0 and 100. Conveniently re-sized, this number can be interpreted as a probability.

Consider the same model space as before, defined in (**??**), with $W_t = X_{t-1}$. For each of these models and for each time t we define $p_{t,j}$, which we call a Google probability:

$$p_{t,j} =_{s \in I^j} Z_{st\, s \in I^{\sim j}} \left(1 - Z_{st}\right).$$

where I^j indicates which variables are in model j. For instance, if model j is the TVP regression model which contains lags of the third and seventh explanatory variables then $I^j = \{3, 7\}$. In a similar fashion, we denote the explanatory variables which are excluded from model j by $I^{\sim j}$. It can be seen that $_{j=1}^{J} p_{t,j} = 1$ and that each Google model probability reflects increases or decreases in internet searches. In our example where $I^j = \{3, 7\}$, $p_{t,j}$ will be large in times when internet searches on terms relating to the third and seventh explanatory variables are unusually high and it will be low when such searches are unusually low.

Our modified version of DMA and DMS with Google model probabilities involves implementing the algorithm of Raftery et al (2010), except with the time varying model probabilities altered to reflect the Google model probabilities as:

$$q_{t|t-1,j} = \omega \frac{q^{\alpha}_{t-1|t-1,j}}{\sum_{l=1}^{J} q^{\alpha}_{t-1|t-1,l}} + (1 - \omega)\, p_{t,j} \tag{8}$$

where ω can be selected by the researcher and $0 \leq \omega \leq 1$. If $\omega = 1$ we are back in conventional DMA or DMS as done by Raftery et al (2010), if $\omega = 0$ then $p_{t,s}$ replaces $q_{t|t-1,s}$ in the algorithm (and, hence, the Google model probabilities are driving model switching). Intermediate values of ω will combine the information in the Google internet searches with the Raftery et al (2010) data-based model probabilities.

4 Nowcasting Using DMS and DMA with Google Model Probabilities

4.1 Overview

In this section, we present evidence on the nowcasting performance of various implementations of DMA and DMS using the data set described in Section 2. For each of the nine variables in Table 1, we carry out a nowcasting exercise using several different approaches most of which are either DMA or DMS using (**??**). In particular, we consider $\omega = 0, \frac{1}{2}, 1$. We also categorize our approaches depending on whether Google variables are used as regressors as in (**??**), used in the DMA model probabilities as in (**??**) or not used at all as in (**??**). We stress that all of our DMA and DMS approaches involve TVP regression models. As benchmarks we also present recursive OLS nowcasts using all of the relevant explanatory variables, recursive nowcasts using an AR(2) model and "No change" nowcasts which use the most recently available observation on the dependent variable as its nowcast.

We use mean squared forecast errors (MSFEs) to evaluate the quality of point forecasts and sums of log predictive likelihoods to evaluate the quality of the predictive densities produced by the various methods. Remember though, that our macroeconomic data is available from January 1973 through July 2012, but the Google data only exists since January 2004. In light of this mismatch in sample span, we estimate all our models in two different ways. First, we simply discard all pre-2004 data for all variables and estimate our models using this relatively short sample. Second, we use data back to 1973 for the macroeconomic variables, but pre-2004 we do not use versions of the models involving the unavailable Google data. For instance, when doing DMA with $\omega = \frac{1}{2}$ we proceed as follows: Pre-2004 we do conventional DMA as implemented in Raftery et al (2010) so that $q_{t|t-1,j}$ is defined using (**??**). As of January 2004, however, $q_{t|t-1,j}$ is defined using (**??**).[5] Results using post-2004 data are given in Table 2 with results using data since 1973 being in Table 3. In the former case, the nowcast evaluation period begins in September 2005, in the latter case in January 2004. In both cases, the nowcast evaluation period ends in July 2012. Our OLS and No change benchmark approaches involve only one model and do not produce predictive likelihoods. Hence, only MSFEs are provided for these benchmarks which are put in the column labelled DMA in the tables.

4.2 Discussion of Empirical Results

With 9 variables, two different forecast metrics and two different sample spans, there are 36 different dimensions in which our approaches can be compared. Not surprisingly, we are not finding one approach which nowcasts best in every case. However, there is a strong tendency to find that DMA and DMS

[5] For the case where the Google variables are included as regressors, we only use post-2004 data.

methods nowcast better than standard benchmarks and there are many cases where the inclusion of Google data improves nowcast performance relative to the comparable approach excluding the Google data. Inclusion of Google data in the form of model probabilities is typically (although not always) the best way of including Google data. It is typically the case that DMS nowcasts better than the comparable DMA algorithm, presumably since the ability of DMS to switch quickly between different parsimonious models helps improve nowcasts. The remainder of this sub-section elaborates on these points, going through one macroeconomic variable at a time.

Inflation. For inflation, we find DMS with $\omega = 0$ or $\omega = \frac{1}{2}$ to produce the best nowcasts, regardless of data span or forecast metric. Note that both of these approaches uses Google probabilities. Doing DMS using Google variables as regressors leads to a worse nowcast performance. For instance, Table 2a shows that doing DMS using Google probabilities yields an MSFE of 19.13 but if DMS is done in the conventional manner using Google variables as regressors, the MSFE is 21.08, which is a fairly substantial deterioration. If the Google variables are simply used as regressors in a recursive OLS exercise, the MSFE deteriorates massively to 37.23. Similar results, where relevant, hold for the predictive likelihoods. In Table 2a, the best MSFE for an OLS benchmark model is 24.22 which also is much worse than DMS using Google probabilities.

Industrial Production: As with inflation, there is strong evidence that DMS leads to nowcast improvements over benchmark OLS methods. However, evidence conflicts on the best way to include Google variables. If we use only the post–2004 data, the MSFEs indicate the Google variables are best used as regressors (along with DMS methods). However, predictive likelihoods indicate that DMS with Google model probabilities nowcasts best. However, if we use data since 1973, MSFEs and predictive likelihoods both indicate that simply doing DMS using the macroeconomic variables nowcasts best. Hence, we are finding strong support for the use of DMS, but a less clear story on how or whether Google variables should be used with DMS.

Unemployment: With the post-2004 data, MSFEs indicate support for our DMS approach using Google probabilities, but predictive likelihoods indicate a preference for using the Google variables as regressors (or not at all). When using the post-1973 sample, predictive likelihoods also indicate support for DMS using Google probabilities. However, MSFEs indicate omitting the Google variables leads to the best nowcasts, with conventional DMS and recursive OLS being the winning approaches according to this metric.

Wage inflation: This is a variable for which MSFE and predictive likelihood results are in accordance. For the post-2004 sample they indicate conventional DMS, using the Google variables as regressors, is to be preferred. However, for the post-1973 sample, they indicate DMS using Google probabilities nowcasts best.

Money: The different measures of nowcast performance and sample spans also lead to a consistent story for money supply growth. In particular, DMS with Google probabilities nowcasts best, although there is some disagreement over whether $\omega = 0$ or $\frac{1}{2}$.

Financial Conditions Index: Using MSFEs, both sample spans indicate that DMS with Google data nowcasts best. Predictive likelihoods, though, show a conflict between whether the Google variables should be used as regressors (post-2004 data) or not included at all (post-1973 data).

Oil Price Inflation: For this variable, both nowcast metrics and data spans indicate DMS with $\omega = 0$ nowcasts best. This is the version of DMS which let the Google model probabilities entirely determine which model is selected at each point in time.

Commodity Price Inflation: Using the post-2004 sample, we find the best performance using DMS with the Google variables being used as regressors. However, using the post-1973 sample we find the approaches including the Google model probabilities (either with $\omega = 0$ or $\frac{1}{2}$) to nowcast best.

Term Spread: Using the smaller post-2004 sample, we are finding that DMS using Google variables as regressors narrowly beats approaches using Google probabilities to be the best nowcasting model. However, in the longer sample, approaches which use the Google probabilities nowcast best. We note also that this is one of the few variables where a benchmark approach does well. In particular, using the post-2004 sample, an AR(2) model nowcasts quite well (although it does not beat our DMS approach).

Table 2a: Nowcast Performance (post-2004 data)				
	LogPL		MSFE	
	DMA	DMS	DMA	DMS
	Inflation			
	Google Variables Not Used			
$\omega = 1$	-236.73	-235.38	24.95	23.28
Rec. OLS	-	-	30.75	-
Rec. AR(2)	-	-	24.22	-
No change	-	-	31.20	-
	Google Variables Used in Model Probs			
$\omega = 0.5$	-239.41	-232.29	24.69	19.35
$\omega = 0$	-239.48	-232.36	24.75	19.13
	Google Variables Used as Regressors			
$\omega = 1$	-237.64	-233.23	26.28	21.08
Rec. OLS	-	-	37.23	-
	Industrial Production			
	Google Variables Not Used			
$\omega = 1$	-289.10	-287.78	107.04	104.46
Rec. OLS	-	-	165.51	-
Rec. AR(2)	-	-	114.13	-
No change	-	-	113.83	-
	Google Variables Used in Model Probs			
$\omega = 0.5$	-291.74	-286.46	116.96	110.12
$\omega = 0$	-291.94	-284.42	117.49	109.74
	Google Variables Used as Regressors			
$\omega = 1$	-288.28	-284.98	102.88	95.90
Rec. OLS	-	-	158.12	-

Table 2b: Nowcast Performance (post-2004 data)				
	LogPL		MSFE	
	DMA	DMS	DMA	DMS
	Unemployment			
	Google Variables Not Used			
$\omega = 1$	-124.10	-123.04	0.033	0.033
Rec. OLS	-	-	0.036	-
Rec. AR(2)	-	-	0.038	-
No change	-	-	5.44	-
	Google Variables Used in Model Probs			
$\omega = 0.5$	-133.75	-127.30	0.032	0.033
$\omega = 0$	-134.41	-128.38	0.032	0.035
	Google Variables Used as Regressors			
$\omega = 1$	-127.91	-123.04	0.034	0.033
Rec. OLS	-	-	0.047	-
	Wage Inflation			
	Google Variables Not Used			
$\omega = 1$	-192.95	-190.10	6.52	5.77
Rec. OLS	-	-	9.78	-
Rec. AR(2)	-	-	6.83	-
No change	-	-	6.15	-
	Google Variables Used in Model Probs			
$\omega = 0.5$	-197.80	-194.11	7.16	5.71
$\omega = 0$	-198.02	-194.49	7.17	5.89
	Google Variables Used as Regressors			
$\omega = 1$	-195.25	-189.50	6.72	5.53
Rec. OLS	-	-	11.48	-

Table 2c: Nowcast Performance (post-2004 data)				
	LogPL		MSFE	
	DMA	DMS	DMA	DMS
	Money			
	Google Variables Not Used			
$\omega = 1$	-245.69	-244.53	30.02	29.71
Rec. OLS	-	-	33.50	-
Rec. AR(2)	-	-	28.99	-
No change	-	-	28.69	-
	Google Variables Used in Model Probs			
$\omega = 0.5$	-249.97	-242.72	29.28	27.34
$\omega = 0$	-250.81	-243.97	29.75	26.07
	Google Variables Used as Regressors			
$\omega = 1$	-247.07	-242.90	31.20	28.12
Rec. OLS	-	-	42.77	-
	FCI			
	Google Variables Not Used			
$\omega = 1$	-53.22	-53.64	0.29	0.29
Rec. OLS	-	-	0.30	-
Rec. AR(2)	-	-	0.32	-
No change	-	-	0.45	-
	Google Variables Used in Model Probs			
$\omega = 0.5$	-58.92	-53.29	0.28	0.21
$\omega = 0$	-59.56	-54.56	0.28	0.21
	Google Variables Used as Regressors			
$\omega = 1$	-55.51	-51.56	0.32	0.26
Rec. OLS	-	-	0.48	-

Table 2d: Nowcast Performance (post-2004 data)				
	LogPL		MSFE	
	DMA	DMS	DMA	DMS
	Oil Price Inflation			
	Google Variables Not Used			
$\omega = 1$	-484.51	-479.54	13,219	10,407
Rec. OLS	-	-	17,465	-
Rec. AR(2)	-	-	11,253	-
No change	-	-	12,185	-
	Google Variables Used in Model Probs			
$\omega = 0.5$	-481.39	-475.00	11,678	8,961
$\omega = 0$	-481.60	-474.71	11,857	8,555
	Google Variables Used as Regressors			
$\omega = 1$	-484.63	-479.72	13,241	10,415
Rec. OLS	-	-	29,333	-
	Commodity Price Inflation			
	Google Variables Not Used			
$\omega = 1$	-429.24	-425.50	3,115	2,706
Rec. OLS	-	-	3,925	-
Rec. AR(2)	-	-	2,950	-
No change	-	-	3,254	-
	Google Variables Used in Model Probs			
$\omega = 0.5$	-429.74	-427.97	3,169	2,964
$\omega = 0$	-429.85	-428.49	3,168	2,986
	Google Variables Used as Regressors			
$\omega = 1$	-429.23	-424.71	3,120	2,635
Rec. OLS	-	-	5,193	-

Table 2e: Nowcast Performance (post-2004 data)				
	LogPL		MSFE	
	DMA	DMS	DMA	DMS
	Term Spread			
	Google Variables Not Used			
$\omega = 1$	-87.68	-86.67	0.072	0.072
Rec. OLS	-	-	0.092	-
Rec. AR(2)	-	-	0.068	-
No change	-	-	1.476	-
	Google Variables Used in Model Probs			
$\omega = 0.5$	-99.42	-91.28	0.069	0.081
$\omega = 0$	-100.53	-93.32	0.069	0.091
	Google Variables Used as Regressors			
$\omega = 1$	-91.44	-86.67	0.068	0.072
Rec. OLS	-	-	0.103	-

Table 3a: Nowcast Performance (data since 1973)				
	LogPL		MSFE	
	DMA	DMS	DMA	DMS
	Inflation			
	Google Variables Not Used			
$\omega = 1$	-293.11	-291.56	20.39	19.39
Rec. OLS	-	-	22.80	-
Rec. AR(2)	-	-	24.16	-
No change	-	-	34.10	-
	Google Variables Used in Model Probs			
$\omega = 0.5$	-293.71	-290.95	20.42	18.74
$\omega = 0$	-293.73	-291.71	20.42	19.09
	Industrial Production			
	Google Variables Not Used			
$\omega = 1$	-363.09	-360.27	94.88	88.38
Rec. OLS	-	-	90.43	-
Rec. AR(2)	-	-	90.38	-
No change	-	-	104.35	-
	Google Variables Used in Model Probs			
$\omega = 0.5$	-362.54	-361.41	94.79	93.11
$\omega = 0$	-362.59	-361.04	94.92	92.61

Table 3b: Nowcast Performance (data since 1973)				
	LogPL		MSFE	
	DMA	DMS	DMA	DMS
	Unemployment			
	Google Variables Not Used			
$\omega = 1$	48.28	50.83	0.027	0.025
Rec. OLS	-	-	0.025	-
Rec. AR(2)	-	-	0.030	-
No change	-	-	4.408	-
	Google Variables Used in Model Probs			
$\omega = 0.5$	45.85	46.19	0.029	0.028
$\omega = 0$	45.51	46.15	0.029	0.028
	Wage Inflation			
	Google Variables Not Used			
$\omega = 1$	-232.99	-229.57	6.06	5.44
Rec. OLS	-	-	9.30	-
Rec. AR(2)	-	-	7.71	-
No change	-	-	10.41	-
	Google Variables Used in Model Probs			
$\omega = 0.5$	-233.69	-230.94	6.16	5.42
$\omega = 0$	-233.77	-230.37	6.13	5.39

Table 3c: Nowcast Performance (post-2004 data)				
	LogPL		MSFE	
	DMA	DMS	DMA	DMS
	Money			
	Google Variables Not Used			
$\omega = 1$	-294.56	-293.57	23.46	22.73
Rec. OLS	-	-	23.02	-
Rec. AR(2)	-	-	23.34	-
No change	-	-	24.16	-
	Google Variables Used in Model Probs			
$\omega = 0.5$	-293.99	-290.76	22.97	20.38
$\omega = 0$	-294.11	-291.24	23.07	20.92
	FCI			
	Google Variables Not Used			
$\omega = 1$	-28.63	-28.02	0.17	0.17
Rec. OLS	-	-	0.18	-
Rec. AR(2)	-	-	0.20	-
No change	-	-	0.36	-
	Google Variables Used in Model Probs			
$\omega = 0.5$	-31.71	-31.17	0.18	0.16
$\omega = 0$	-31.70	-31.15	0.18	0.16

Table 3d: Nowcast Performance (post-2004 data)				
	LogPL		MSFE	
	DMA	DMS	DMA	DMS
	Oil Price Inflation			
	Google Variables Not Used			
$\omega = 1$	-610.18	-608.20	10,443	9,836
Rec. OLS	-	-	10,468	-
Rec. AR(2)	-	-	9,740	-
No change	-	-	10,957	-
	Google Variables Used in Model Probs			
$\omega = 0.5$	-609.54	-607.24	10,210	9,269
$\omega = 0$	-609.63	-606.73	10,230	9,064
	Commodity Price Inflation			
	Google Variables Not Used			
$\omega = 1$	-531.01	-528.20	2,230	2,080
Rec. OLS	-	-	2,198	-
Rec. AR(2)	-	-	2,200	-
No change	-	-	2,710	-
	Google Variables Used in Model Probs			
$\omega = 0.5$	-529.57	-528.19	2,230	2,079
$\omega = 0$	-529.64	-527.63	2,235	2,084

Table 3e: Nowcast Performance (post-2004 data)				
	LogPL		MSFE	
	DMA	DMS	DMA	DMS
	Term Spread			
	Google Variables Not Used			
$\omega = 1$	2.980	6.239	0.062	0.056
Rec. OLS	-	-	0.109	-
Rec. AR(2)	-	-	0.083	-
No change	-	-	1.382	-
	Google Variables Used in Model Probs			
$\omega = 0.5$	2.374	6.785	0.062	0.053
$\omega = 0$	2.484	6.754	0.061	0.052

5 Further Discussion and Conclusions

The preceding discussion reveals a wide variety of findings. The following main conclusions emerge:

- First, the inclusion of Google data leads to sizeable improvements in nowcast performance. This result complements the existing literature by showing that Google search variables are not only useful when dealing with specific disaggregate variables, but can be used to improve nowcasting of broad macroeconomic aggregates.

- Second, and despite the crude procedure we adopted to create the Google variables, we also find that it is often (albeit not invariably) the case that the information in the Google variables is best included in the form of model probabilities as opposed to simply including Google variables as regressors. The intuition that Google search volumes may provide the econometrician with useful information about which variable is important at each point in time opens the way to a new and more extensive use of this vast database.

- Finally, Google probabilities make sense in a context where the economy is not constant, and are therefore particularly suited to deal with the recent crisis. However, their potential must be exploited with opportune techniques allowing for model change and parsimony. We compared different techniques responding to such requirements. DMS proved to be a particularly good method for improving nowcast performance in the models we are dealing with, leading to substantial improvements over common benchmarks. It is also worth noting that DMS is a strategy which often nowcasts best, but even when it does not it does not go too far wrong. Our simple benchmarks, using OLS methods, sometimes also provide reasonable nowcasts but occasionally produce very bad nowcasts.

This is a first and so far successful attempt to use Google variables to improve macroeconomic nowcasting. We proposed two different uses of these variables, one of which, to our knowledge, completely new and close to the spirit ("what are people concerned about?") in which these variables are collected. Additional research will be needed to make these results more robust. Our construction of the Google variables, in particular, is extremely simple, and it is not unlikely that a more accurate choice in the searches or a different method of averaging may lead to further improvements in their use.

References

Artola, C. and Galan, E. (2012). "Tracking the future on the web: Construction of leading indicators using internet searches," Documentos Ocasionales No. 1203, Bank
of Spain.

Askitas, N. and Zimmermann, K. (2009). "Google econometrics and unemployment nowcasting," DIW Berlin, Discussion Paper 899.

Canova, F. (1993). "Modelling and nowcasting exchange rates using a Bayesian time varying coefficient model," *Journal of Economic Dynamics and Control*, 17, 233-262.

Canova, F. and Ciccarelli, M. (2009). "Estimating multi-country VAR models," *International Economic Review*, 50, 929-959.

Canova, F. and Gambetti, L. (2009). "Structural changes in the US economy: Is there a role for monetary policy?" *Journal of Economic Dynamics and Control*, 33, 477-490.

Carriere-Swallow, Y. and Labbe, F. (2011). "Nowcasting with Google trends in an emerging market," *Journal of nowcasting*, doi: 10.1002/for.1252.

Chamberlin, G. (2010). "Googling the present," U.K. Office for National Statistics Economic & Labour Market Review, December 2010.

Chan, J., Koop, G., Leon-Gonzalez, R. and Strachan, R. (2012). "Time varying dimension models," Journal of Business and Economic Statistics, 30, 358-367.

Choi, H. and Varian, R. (2009). "Predicting initial claims for unemployment insurance using Google Trends," Google Technical Report.

Choi, H. and Varian, R. (2011). "Predicting the present with Google Trends," Google Technical Report.

Cogley, T. and Sargent, T. (2001). "Evolving post-World War II inflation dynamics," *NBER Macroeconomic Annual*, 16, 331-373.

Cogley, T. and Sargent, T. (2005). "Drifts and volatilities: Monetary policies and outcomes in the post WWII U.S," *Review of Economic Dynamics*, 8, 262-302.

D'Amuri, F. and Marcucci, J. (2009). "'Google it!' nowcasting the US unemployment rate with a Google job search index," Institute for Economic and Social Research Discussion Paper 2009-32.

Dangl, T. and Halling, M. (2012). "Predictive regressions with time varying coefficients," *Journal of Financial Economics,* forthcoming.

Giannone, D., Lenza, M., Momferatou, D. and Onorante, L. (2010). "Short-term inflation projections: A Bayesian vector autoregressive approach," ECARES working paper 2010-011.

Hellerstein, R. and Middeldorp, M. (2012). "Nowcasting with internet search data," Liberty Street Economics Blog of the Federal Reserve Bank of New York, January 4, 2012.

Kholodilin, K., Podstawski, M. and Siliverstovs, S. (2010). "Do Google searches help in nowcasting private consumption? A real-time evidence for the US," KOF Swiss Economic Institute Discussion Paper No. 256.

Koop, G. and Korobilis, D. (2012). "Forecasting inflation using dynamic model averaging," International Economic Review, 53, 867-886.

Koop, G., Leon-Gonzalez, R. and Strachan, R. (2009). "On the evolution of the monetary policy transmission mechanism," *Journal of Economic Dynamics and Control*, 33, 997–1017.

Koop, G. and Onorante, L. (2012). "Estimating Phillips curves in turbulent times using the ECB's survey of professional nowcasters," European Central Bank, working paper number 1422.

Koop, G. and Tole, L. (2013). "Forecasting the European carbon market," Journal of the Royal Statistical Society, Series A, 176, Part 3, pp. 723-741.

McLaren, N. and Shanbhoge, R. (2011). "Using internet search data as economic indicators," Bank of England Quarterly Bulletin, June 2011.

Nicoletti, G. and Passaro, R. (2012). "Sometimes it helps: Evolving predictive power of spreads on GDP," European Central Bank, working paper number 1447.

Primiceri. G., (2005). "Time varying structural vector autoregressions and monetary policy," *Review of Economic Studies*, 72, 821-852.

Raftery, A., Karny, M. and Ettler, P. (2010). "Online prediction under model uncertainty via dynamic model averaging: Application to a cold rolling mill," *Technometrics*, 52, 52-66.

RiskMetrics (1996). *Technical Document* (Fourth Edition). Available at http://www. riskmetrics.com/system/files/private/td4e.pdf.

Scott, S. and Varian, H. (2012). "Bayesian variable selection for nowcasting economic time series," manuscript available at

http://people.ischool.berkeley.edu/~hal/people/hal/papers.html.

Schmidt, T. and Vosen, S. (2009). "Nowcasting private consumption: Survey-based indicators vs. Google trends," Ruhr Economic Papers, No. 155.

Stock, J. and Watson, M. (1996). "Evidence on structural instability in macroeconomic time series relations," *Journal of Business and Economic Statistics*,14, 11-30.

Suhoy, T. (2009). "Query indices and a 2008 downturn: Israeli data," Bank of Israel Discussion Paper No. 2009.06.

West, M. and Harrison, J. (1997). *Bayesian nowcasting and Dynamic Models*, second edition, New York: Springer.

Wu, L. and Brynjolfsson, E. (2010). "The future of prediction: How Google searches foreshadow housing prices and sales," MIT Technical Report.

Appendix: Google Search Terms

The following tables list the google search terms we use along with the macroeconomic variable each corresponds to.

Table A1: Categorization of Google Search Terms	
Google search term	**Category**
steel price	Commodity Price Inflation
food price	Commodity Price Inflation
copper price	Commodity Price Inflation
stock compensation	Financial Conditions Index
investment banking	Financial Conditions Index
growth equity	Financial Conditions Index
goldman sachs	Financial Conditions Index
equity compensation	Financial Conditions Index
us gdp growth	Industrial Production
urban growth	Industrial Production
the great depression	Industrial Production
tax calculator	Industrial Production
small business growth	Industrial Production
sales growth	Industrial Production
sales compensation	Industrial Production
revenue growth	Industrial Production
recession	Industrial Production
recession inflation	Industrial Production
market growth	Industrial Production
growth	Industrial Production
growth industries	Industrial Production
growth financial	Industrial Production
growth company	Industrial Production
growth companies	Industrial Production
great depression	Industrial Production
great depression deflation	Industrial Production
gdp growth	Industrial Production
economy	Industrial Production

Table A2: Categorization of Google Search Terms	
Google search term	**Category**
economic growth	Industrial Production
cycle	Industrial Production
crisis	Industrial Production
business growth	Industrial Production
business cycle	Industrial Production
what is inflation	Inflation
what is deflation	Inflation
us inflation	Inflation
us inflation rates	Inflation
us inflation rate	Inflation
us inflation index	Inflation
us deflation	Inflation
united states inflation	Inflation
u.s. inflation	Inflation
real inflation	Inflation
rate of inflation	Inflation
price inflation	Inflation
price index	Inflation
national inflation	Inflation
investing deflation	Inflation
inflation	Inflation
inflation usa	Inflation
inflation stocks	Inflation
inflation rates	Inflation
inflation rate	Inflation
inflation or deflation	Inflation
inflation money	Inflation
inflation index	Inflation

Table A3: Categorization of Google Search Terms	
Google search term	**Category**
inflation in us	Inflation
inflation graph	Inflation
inflation forecast	Inflation
inflation deflation	Inflation
inflation definition	Inflation
inflation data	Inflation
inflation chart	Inflation
inflation calculator	Inflation
inflation and deflation	Inflation
india inflation	Inflation
historical inflation	Inflation
high inflation	Inflation
fed deflation	Inflation
economic inflation	Inflation
economic deflation	Inflation
depression deflation	Inflation
deflation	Inflation
deflation rate	Inflation
deflation interest rates	Inflation
deflation in us	Inflation
deflation gold	Inflation
deflation economy	Inflation
definition inflation	Inflation
definition deflation	Inflation
define inflation	Inflation
debt deflation	Inflation
current inflation	Inflation
current inflation rate	Inflation

Table A4: Categorization of Google Search Terms	
Google search term	**Category**
cpi	Inflation
cpi index	Inflation
cost of inflation	Inflation
consumer price index	Inflation
money	Money Supply
money deflation	Money Supply
monetary policy	Money Supply
monetary deflation	Money Supply
oil production	Oil Price Inflation
oil prices	Oil Price Inflation
oil price	Oil Price Inflation
gasoline price	Oil Price Inflation
gas price	Oil Price Inflation
energy production	Oil Price Inflation
energy price	Oil Price Inflation
electricity price	Oil Price Inflation
diesel price	Oil Price Inflation
production	Industrial Production
production jobs	Industrial Production
production company	Industrial Production
production companies	Industrial Production
us interest rate	Term Spread
the fed	Term Spread
real interest rate	Term Spread
prime rate	Term Spread
prime interest rate	Term Spread
mortgage rate	Term Spread
mortgage interest rates	Term Spread

Table A5: Categorization of Google Search Terms	
Google search term	**Category**
lower interest rate	Term Spread
libor	Term Spread
libor rate	Term Spread
libor interest rate	Term Spread
interest rates	Term Spread
interest rates inflation	Term Spread
interest rate	Term Spread
interest rate trends	Term Spread
interest rate risk	Term Spread
interest rate reduction	Term Spread
interest rate predictions	Term Spread
interest rate news	Term Spread
interest rate mortgage	Term Spread
interest rate model	Term Spread
interest rate inflation	Term Spread
interest rate history	Term Spread
interest rate forecast	Term Spread
interest rate fed	Term Spread
interest rate drop	Term Spread
interest rate cuts	Term Spread
interest rate cut	Term Spread
interest rate chart	Term Spread
interest rate calculator	Term Spread
feds interest rate	Term Spread
federal reserve	Term Spread
federal interest rate	Term Spread
fed	Term Spread
fed rates	Term Spread

Table A6: Categorization of Google Search Terms	
Google search term	**Category**
fed rate	Term Spread
fed rate cut	Term Spread
fed interest rates	Term Spread
fed interest rate	Term Spread
fed cut	Term Spread
discount rate	Term Spread
current interest rate	Term Spread
washington unemployment	Unemployment Rate
us unemployment	Unemployment Rate
us unemployment rate	Unemployment Rate
unemployment	Unemployment Rate
unemployment statistics	Unemployment Rate
unemployment rates	Unemployment Rate
unemployment rate	Unemployment Rate
unemployment pa	Unemployment Rate
unemployment office	Unemployment Rate
unemployment michigan	Unemployment Rate
unemployment insurance	Unemployment Rate
unemployment great depression	Unemployment Rate
unemployment extension	Unemployment Rate
unemployment depression	Unemployment Rate
unemployment checks	Unemployment Rate
unemployment check	Unemployment Rate
unemployment benefits	Unemployment Rate
texas unemployment	Unemployment Rate
subsidies	Unemployment Rate
state compensation fund	Unemployment Rate
oregon unemployment	Unemployment Rate

Table A7: Categorization of Google Search Terms

Google search term	**Category**
ohio unemployment	Unemployment Rate
ny unemployment	Unemployment Rate
nj unemployment	Unemployment Rate
new york unemployment	Unemployment Rate
michigan works	Unemployment Rate
michigan works unemployment	Unemployment Rate
michigan state unemployment	Unemployment Rate
marvin unemployment	Unemployment Rate
marvin michigan unemployment	Unemployment Rate
job growth	Unemployment Rate
florida unemployment	Unemployment Rate
federal unemployment	Unemployment Rate
employee benefits	Unemployment Rate
depression unemployment rate	Unemployment Rate
compensation packages	Unemployment Rate
compensation package	Unemployment Rate
california unemployment	Unemployment Rate
workers compensation	Wage Inflation
workers compensation ohio	Wage Inflation
workers compensation insurance	Wage Inflation
what is compensation	Wage Inflation
walmart wages	Wage Inflation
wages	Wage Inflation
wages calculator	Wage Inflation
wage	Wage Inflation
wage inflation	Wage Inflation
vice president salary	Wage Inflation
us wages	Wage Inflation

#

Table A8: Categorization of Google Search Terms	
Google search term	**Category**
unpaid wages	Wage Inflation
union wages	Wage Inflation
total compensation	Wage Inflation
state wages	Wage Inflation
state employee wages	Wage Inflation
salary	Wage Inflation
salary tax calculator	Wage Inflation
salary survey	Wage Inflation
salary schedule	Wage Inflation
salary requirements	Wage Inflation
salary raise	Wage Inflation
salary grade	Wage Inflation
salary comparison	Wage Inflation
salary calculator hourly	Wage Inflation
salaries	Wage Inflation
real wages	Wage Inflation
project manager salary	Wage Inflation
pilot salary	Wage Inflation
paycheck calculator	Wage Inflation
nfl salary	Wage Inflation
nfl minimum salary	Wage Inflation
minimum wages	Wage Inflation
labor wages	Wage Inflation
labor and wages	Wage Inflation
job wages	Wage Inflation
investment banking salary	Wage Inflation
incentive compensation	Wage Inflation
human resources salary	Wage Inflation

Table A9: Categorization of Google Search Terms	
Google search term	**Category**
human resources compensation	Wage Inflation
hr compensation	Wage Inflation
hourly wages	Wage Inflation
gross wages	Wage Inflation
gross salary	Wage Inflation
federal wages	Wage Inflation
federal salary	Wage Inflation
executive compensation	Wage Inflation
employment wages	Wage Inflation
employee wages	Wage Inflation
employee compensation	Wage Inflation
director compensation	Wage Inflation
deferred compensation	Wage Inflation
compensation	Wage Inflation
compensation time	Wage Inflation
compensation system	Wage Inflation
compensation structure	Wage Inflation
compensation resources	Wage Inflation
compensation plans	Wage Inflation
compensation plan	Wage Inflation
compensation manager	Wage Inflation
compensation consulting	Wage Inflation
compensation analyst	Wage Inflation
china wages	Wage Inflation
ceo salary	Wage Inflation
ceo compensation	Wage Inflation
calculate salary	Wage Inflation
bonus compensation	Wage Inflation
benefits and compensation	Wage Inflation
average wages	Wage Inflation
average salary	Wage Inflation
average nfl salary	Wage Inflation
annual compensation	Wage Inflation